A Review of Psychological Assessment Instruments for Use in Bariatric Surgery Evaluations

Ryan J. Marek
Kent State University

Leslie J. Heinberg
Cleveland Clinic Lerner College of Medicine

Megan Lavery
Cleveland Clinic Bariatric and Metabolic Institute, Cleveland, Ohio

Julie Merrell Rish and Kathleen Ashton
Cleveland Clinic Lerner College of Medicine

Bariatric surgery is a viable treatment option for patients with extreme obesity and associated medical comorbidities; however, optimal surgical outcomes are not universal. Surgical societies, such as the American Society for Metabolic and Bariatric Surgery (ASMBBS), recommend that patients undergo a presurgical psychological evaluation that includes reviewing patients’ medical charts, conducting a comprehensive clinical interview, and employing some form of objective psychometric testing. Despite numerous societies recommending the inclusion of self-report assessments, only about 2/3 of clinics actively use psychological testing—some of which have limited empirical support to justify their use. This review aims to critically evaluate the psychometric properties of self-report measures when used in bariatric surgery settings and provide recommendations to help guide clinicians in selecting instruments to use in bariatric surgery evaluations. Recommended assessment batteries include use of a broadband instrument along with a narrowband eating measure. Suggestions for self-report measures to include in a presurgical psychological evaluation in bariatric surgery settings are also provided.

Keywords: assessment, evaluations, bariatric, obesity, psychology

Obesity is prevalent and has become a public health crisis due to its association with a number of physical health concerns. Approximately 1.4 billion adult individuals worldwide are considered overweight (WHO, 2013). Of those, nearly 600 million men and women were considered obese—defined as having a body mass index (BMI) of 30 kg/m² or more. In the United States, approximately 6% of the population have a BMI ≥40 kg/m² (Flegal, Carroll, Ogden, & Curtin, 2010). Specifically, obesity is associated with cardiovascular risk factors, Type II Diabetes, obstructive sleep apnea, and having an increased risk of developing cancer (Calle, Rodriguez, Walker-Thurmond, & Thun, 2003; Mechanick et al., 2008, 2013; Strobel & Rosen, 1996; von Eyben et al., 2003). Early mortality is also associated with obesity (WHO, 2013). Importantly, obesity is also preventable and treatable.

Because obesity is associated with a number of medical problems, long-term weight loss can be difficult. For individuals with a BMI ≥40 kg/m² or BMI ≥35 kg/m² with a comorbid medical condition, bariatric surgery is recommended as a treatment option for obesity. Bariatric surgery has demonstrated better sustained weight loss outcomes than diet, lifestyle changes, or pharmacotherapy (Brethauer, Chand, & Schauer, 2006; Courcoulas et al., 2013; Schauer et al., 2014). One year following most bariatric surgery procedures (such as the Roux-en-Y Gastric Bypass or Sleeve Gastrectomy), patients can expect to lose 50% or more of their excess weight (Boza et al., 2010; Mitchell, Garcia, de Zwaan, & Horbach, 2012). Moreover, a majority of patients maintain their weight loss up to 10 years following the procedure (Valezi et al., 2013). In addition to weight loss, bariatric surgery is also associated with a significant reduction (and sometimes a complete resolution) of associated medical conditions (Deitel, 2002; Gami, Caples, & Somers, 2003; Mechanick et al., 2013). Beyond improving medical conditions, bariatric surgery is also associated with a decrease in all-cause mortality (Sjostrom, 2008). Despite bariatric surgery being associated with dramatic improvements in weight loss and medical comorbidities, a subsample of patients (range: 15%–50%) experience poor outcomes in the form of limited initial weight loss or weight regain (Magro et al., 2008; Mechanick et al., 2013; O’Brien, Dixon, & Brown, 2004; Odom et al., 2010; Sjostrom et al., 2007). Behavioral lifestyle changes are necessary to maintain results. Other suboptimal outcomes include alcohol abuse (King et al., 2012) or engaging in poor eating habits (Conceição et al., 2014; de Zwaan et al., 2010). Moreover, there is a high prevalence of depression-related disorders before and after bariatric surgery (Mitchell et al., 2014) as well as an...
elevated risk for suicide following the procedure (Tindle et al., 2010).

Because presurgical psychological factors have been associated with weight regain and recurrence of behavioral problems after surgery (Conceição et al., 2014; de Zwaan et al., 2011; de Zwaan et al., 2010; King et al., 2012), the American Association of Clinical Endocrinologists (AACE), the American Society for Metabolic and Bariatric Surgery (ASMBS), and The Obesity Society (TOS) Guidelines for the Clinical Management of Bariatric Surgery Patients in addition to the guidelines set forth by the National Institute of Health (NIH) have long recommended that assessment of bariatric surgery candidates should include presurgical psychological evaluations (Mechanick et al., 2008, 2013; NIH, 1991). Specifically, the goals of presurgical psychological evaluations are to identify and treat preexisting psychopathology prior to surgery, identify patients who may need additional postoperative care, and identify alternative treatment strategies if the patient is not deemed a candidate for the procedure they are seeking (Block & Sarwer, 2013). Presurgical psychological evaluations should be composed of a medical chart review, a comprehensive clinical interview, and some form of objective psychological testing. Although there are some psychological contraindications to surgery (e.g., current substance abuse/dependence, uncontrolled psychotic symptoms, etc.), psychological evaluations should be used to enhance surgical outcomes. Presurgical evaluations should also be a part of a multidisciplinary team approach to understand patients' benefits and risks of proceeding with surgery.

Medical chart reviews and comprehensive psychodiagnostic interviews are universally being used across bariatric settings (Bauchowitz et al., 2005; Fabricatore, Crerand, Wadden, Sarwer, & Krasucki, 2006). Domains recommended to be assessed during the evaluation include eating pattern/weight loss attempt history, developmental/family history, current/past mental health treatment, substance use history, cognition (i.e., ability to provide informed consent, understanding of the surgical procedures, risks, etc.), social support, current coping skills, adherence, and motivations/expectations for surgery (LeMont, Moorehead, Parish, Reto, & Ritz, 2004; Mechanick et al., 2013). Templates for conducting the clinical interview that assess the aforementioned domains have been published (Sogg & Mori, 2008/2009; Wadden & Foster, 2006). The Weight and Lifestyle Inventory (WALI; Wadden & Foster, 2006) is a questionnaire designed to assess patient’s weight and dieting histories, activity habits, social and psychological status, and current life stressors. It yields moderate to good test-retest reliability coefficients (Wadden et al., 2006), yields a 60–90 min administration time, and can be used to guide the clinical interview (Wadden & Sarwer, 2006). Wadden and Sarwer (2006) also provide insight on how educating the patients about the surgery and helping them chose which treatment option is best for them likely offer incremental benefits beyond the interview itself. The Boston Interview (Sogg & Mori, 2008) was developed as a clinical interview template designed to assess the following domains: weight, diet, and nutrition history, eating pathology, medical history, understanding of surgical risks/benefits, motivations for surgery, interpersonal functioning, and a psychiatric screener. To date, no empirical evidence (e.g., interrater reliability, predictive validity) has been published on the Boston Interview.

Although clinical interviews are universally used in this setting, only 50%—66% of bariatric surgery clinics include an objective psychological instrument as part of their evaluation battery despite recommendations to use standardized self-report measures (Bauchowitz et al., 2005; Fabricatore et al., 2006). Psychological testing also allows clinicians to gather a large amount of information quickly and synthesize this information, thus maximizing use of professional time. Using objective psychometric testing in conjunction with a clinical interview offers more precise, empirically validated information about the patient than a clinical interview alone (Grove & Meehl, 1996). Including objective psychological assessment in bariatric surgery evaluations can provide clinical information on risk factors that the patient may be sensitive to disclosing during an interview, aid or challenge differential diagnoses, and provide information on the extent to which a candidate is over- or underreporting symptoms (LeMont et al., 2004).

A majority of clinics who use an objective psychological instrument report using either the broadband Minnesota Multiphasic Personality Inventory – 2 (MMPI-2; Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2001) or the Beck Depression Inventory–II (BDI-II; Beck, Steer, & Brown, 1996). Use of instruments varies considerably across clinics, such that some clinics regularly use other broadband instruments such as the Personality Assessment Inventory (PAI; Morey, 1996) and the Million Behavioral Medicine Diagnosis (MBMD), and others use solely narrowband instruments such as the Eating Disorder Examination Questionnaire (EDE-Q; Fairburn & Cooper, 1993) or the Alcohol Use Disorder Identification Test (AUDIT; Saunders, Aasland, Babor, De la Fuente, & Grant, 1993).

To date, no studies have critically reviewed the psychometric properties of objective psychological instruments when used with bariatric surgery settings. Moreover, despite only one study reviewing instrument content (LeMont et al., 2004), no updated clinical guidelines exist to aid clinicians in designing a well-balanced assessment battery for bariatric surgery evaluations.1 The purpose of this review is to critically examine the psychometric properties of self-report instruments being used in bariatric surgery clinics. Instruments were selected for review based on previously published survey data suggesting that more than 5% of bariatric clinics endorsed using the measure (Bauchowitz et al., 2005; Fabricatore et al., 2006; Wallfish, Vance, & Fabricatore, 2007).

Psychological domains for assessment are reviewed first. Afterword, each psychological instrument being used in the field is described in detail, and the literature that reports psychometric data for that specific instrument is reviewed. Both broadband instruments and narrowband instruments that are currently being used in bariatric surgery clinics according to published survey data are discussed (Bauchowitz et al., 2005; Fabricatore et al., 2006; Wallfish et al., 2007). Newer instruments that may have potential utility for being used in this population are highlighted, but have limited to no research in bariatric surgery settings available (e.g., Loss of Control Over Eating Scale). Recommendations regarding assessment batteries with the inclusion of broadband and narrowband instruments is over- or underreporting symptoms (Lemont et al., 2004).

1 Although LeMont et al.’s (2004) article is still available to download at the time this review was written (https://asmbs.org/resources/pre-surgical-psychological-assessment). ASMBS has placed a disclosure statement in 2011 both on the web site and in the document that state the document has never been updated, is currently being substantially revised, and guards readers that some suggestions may no longer be relevant or supported.
band instruments that yield strong psychometric properties are made based on current published literature in the area.

Health-related quality of life (HRQOL) measures and cognitive assessment instruments are beyond the scope of this review. For a summary of these assessments, see Anderson et al.’s (2015) recent systematic review of prospective, long-term studies examining HRQOL in bariatric surgery samples and see Gunstad, Mueller, Stanek, and Spitznagel (2012) for a review of cognitive assessments and example neuropsychiatric batteries to use in bariatric surgery settings.

Psychological Domains Recommended for Assessment

Bariatric patients have a higher prevalence of psychological disorders than the general population (Kalarchian et al., 2007; Mitchell, Selzer, et al., 2012). As discussed earlier, psychosocial factors are associated with weight regain and recurrence of behavioral problems after surgery in some studies, though this is not a consistent trend (Livhits et al., 2012). Studies that support psychopathology predicting diminished outcomes suggest that internalizing dysfunction (mood/anxiety-related), disorders associated with poor eating habits, externalizing dysfunctions (impulsive/substance-abuse behaviors), and poor cognitive functioning are related to suboptimal outcomes (Cole, Dixon, & O’Brien, 2008; Heneghan, Heinberg, Windover, Rogula, & Schauer, 2012; Sarwer, Allison, Bailer, Faulconbridge, & Wadden, 2013; Spitznagel et al., 2014; Suzuki, Haimovici, & Chang, 2012). The following section briefly reviews why a broad assessment of these psychological domains should be considered. For more thorough reviews of the literature, please refer to an integrated review by Marek, Ben-Porath, and Heinberg (2016).

Internalizing Psychopathology

Depression- and anxiety-related disorders are quite prevalent in presurgical bariatric surgery patients (de Zwaan et al., 2011; Kalarchian et al., 2007; Mitchell, Selzer, et al., 2012). As many as 15.6% meet diagnostic criteria for a current depression-related disorder and 24% meet criteria for a current anxiety-related disorder. Lifetime prevalence of depression and anxiety disorders are as high as 45.5% and 37.5%, respectively. Although the prevalence for current mood-related disorders tends to decrease shortly after surgery, it is suggested that patients begin to report more moderate symptoms of depression between the 1- and 3-year postoperative follow-up (Mitchell, King, et al., 2014). As noted earlier, risk for suicide is also a concern after surgery (Tindle et al., 2010). Moreover, antidepressant medication is inadequately absorbed due to anatomical changes in the digestive system (Roerig et al., 2012; Roerig et al., 2013). In a review of the literature, Heneghan, Heinberg, Windover, Rogula, and Schauer (2012) reported that the association between obesity and suicide risk persists after treatment for obesity. Studies examining the relationship between presurgical depression and short-term outcomes (1 to 3 years postsurgery) have produced mixed results. For example, some studies support mood- and anxiety-related measures predict suboptimal weight loss (de Zwaan et al., 2011; Kalarchian et al., 2008; Semanscin-Doerr, Windover, Ashton, & Heinberg, 2010). On the other hand, other studies have found no associations between weight loss outcomes and measures of mood and anxiety (Marek et al., 2015; Tarescavage, Wygant, Boutacoff, & Ben-Porath, 2013). Conversely, some studies have reported associations that demonstrate presurgical mood and anxiety are correlated with better weight loss outcomes (Averbukh et al., 2003; Odom et al., 2010). Research looking at outcomes beyond weight loss, such as physical activity, development of body image concerns, or adherence to nutrition guidelines, is limited.

Eating-Related Behaviors

Disordered eating and poor eating-related behaviors are commonly observed in this population. The prevalence rate for binge eating disorder as outlined in the DSM–IV–TR and DSM–5 ranges from 4.2% to 27.1% (Allison, Wadden et al., 2006; Kalarchian et al., 2007; Marek, Ben-Porath, Ashton, & Heinberg, 2014a; Mitchell, Selzer, et al., 2012). Night eating syndrome (described as awakening from sleep to eat or engaging in excessive food consumption after the last meal of the day) and graze eating (continuous snacking over a long period of time) are also considered problematic in the bariatric population (Allison, Goel, & Ahima, 2014; Allison, Wadden et al., 2006). Similar to mood and anxiety, the relationship between presurgical eating disordered behavior and weight loss in this population is inconsistent, with some studies suggesting eating disorders are associated with suboptimal weight loss outcomes (Green, Dymek-Valentine, Pyduk, le Grange, & Alverdy, 2004; Hsu, Sullivan, & Benotti, 1997), and others suggesting there is no association between presurgical eating disorders and weight loss outcomes (Busetto et al., 2005; Conceição et al., 2014; White, Masheb, Rothschild, Burke-Martindale, & Grilo, 2006). Notably, studies that have used external criteria in addition to weight loss outcomes have found support that patients who binge eat, graze eat, or report a sense of “loss of control” over eating prior to surgery are at a higher risk for engaging in graze eating and a sense of “loss of control” over eating after surgery (Conceição et al., 2014; Saunders, 2004).

Externalizing Psychopathology

Alcohol use, substance use, and problems associated with impulse control are also considered contraindications of surgery if left unaddressed and untreated. In the bariatric surgery population, the lifetime prevalence of substance use disorders is more than twice as high as the general population (~33.2%; Mitchell, Selzer, et al., 2012). Similar to antidepressants, the pharmacokinetics of alcohol are also altered (Woodard, Downey, Hernandez-Boussard, & Morton, 2011). Pharmacokinetic changes following some bariatric surgery procedures further accelerate alcohol absorption, making postsurgical risk of alcohol misuse problematic in bariatric surgery patients. For example, a 5 oz. glass of red wine consumed after a Roux-en-Y Gastric Bypass will raise blood alcohol content to over three times the rate of what is typically observed had bariatric surgery not occurred. Pre- and postsurgical marijuana use is also associated with a greater likelihood of engaging in poorer eating habits postoperatively (Vidot et al., 2015). Absorption and the effects of other substances, such as stimulants, are unknown and warrant further research. In regards to outcomes, the literature is largely inconsistent in relation to this domain as well. Some studies suggest a history of alcohol or drug are associated with
better short-term weight loss outcomes (Clark et al., 2003; Dixon, Dixon, & O’Brien, 2001; Heinberg & Ashton, 2010) and some suggest no association with weight loss outcomes or health-related quality of life (Black, Goldstein, & Mason, 2003; Kopec-Schrader, Gertler, Ramsey-Stewart, & Beumont, 1994; Sears, Fillmore, Bui, & Rodriguez, 2008; Suzuki et al., 2012). Clark et al. (2003) suggest that the association between a history of substance use and weight loss outcomes are mediated by a history of treatment for substance abuse therefore, a history of substance abuse treatment may serve as a good prognostic factor.

Thought Disorder

Untreated or unstable thought disorder (i.e., schizophrenia) is generally considered a contraindication to weight loss surgery (Gross & van Elst, 2012); however, few studies have examined whether bariatric surgery is safe and effective for patients diagnosed with thought disorders in stable treatment. Two empirical studies examining whether patients with well-controlled thought disorder differed in weight loss outcomes and complications after surgery, reporting minimal to no differences in complication rates or weight loss outcomes between patients with thought disorders and those with no reported psychosocial complications (Hamoui, Kingsbury, Anthone, & Crookes, 2004; Shelby, Labott, & Stout, 2015). Other potential concerns around the long-term efficacy for patients with formal thought disorder include weight regain from antipsychotic medication, effects from anesthesia (such as prolonged delirium), adherence and understanding of the procedure, complications around their mental health care due to anatomy changes, and deficits associated with neurocognitive domains such as attention, executive functioning, and memory (Allison et al., 1999; Kudoh, Takase, Takahira, Katagai, & Takazawa, 2003; Spitznagel et al., 2014; Zygmun, Olfsen, Boyer, & Mechanic, 2002). Therefore, it is important to properly identify and monitor patients with features associated with thought disorders to assure safety and efficacy of the surgical treatment.

Table 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>Internal consistency</th>
<th>Validity with external criteria</th>
<th>Normative data</th>
<th>Validity scales</th>
<th>Administration time</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadband measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMPI-2</td>
<td>.34–.87</td>
<td>WL</td>
<td>Yes</td>
<td>Yes</td>
<td>90–120 min</td>
<td>Yes</td>
</tr>
<tr>
<td>MMPI-2-RF</td>
<td>M = .74</td>
<td>WL, ADH, PSY, EB, DX</td>
<td>Yes</td>
<td>Yes</td>
<td>35–50 min</td>
<td>Yes</td>
</tr>
<tr>
<td>PAI</td>
<td>M = .78</td>
<td>PSY</td>
<td>Yes</td>
<td>Yes</td>
<td>50–60 min</td>
<td>Yes</td>
</tr>
<tr>
<td>SCL-90-R</td>
<td>M = .84</td>
<td>ADH, PSY</td>
<td>Yes</td>
<td>No</td>
<td>12–15 min</td>
<td>Yes</td>
</tr>
<tr>
<td>MBMD</td>
<td>Median = .70</td>
<td>Included in Test Manual</td>
<td>Yes</td>
<td>Yes</td>
<td>20–25 min</td>
<td>Yes</td>
</tr>
<tr>
<td>MCMI-III</td>
<td>NA</td>
<td>WL, PSY</td>
<td>No</td>
<td>Yes</td>
<td>25–30 min</td>
<td>Yes</td>
</tr>
<tr>
<td>BPI</td>
<td>NA</td>
<td>WL</td>
<td>No</td>
<td>Yes</td>
<td>35 min</td>
<td>Yes</td>
</tr>
<tr>
<td>Narrowband measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression/anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI-II</td>
<td>.91</td>
<td>DX, PSY</td>
<td>Yes</td>
<td>No</td>
<td>5–10 min</td>
<td>Yes</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>.86–.89</td>
<td>DX, PSY</td>
<td>Yes</td>
<td>No</td>
<td>5 min</td>
<td>No</td>
</tr>
<tr>
<td>CES-D</td>
<td>.85–.94</td>
<td>DX, PSY</td>
<td>No</td>
<td>No</td>
<td>10 min</td>
<td>No</td>
</tr>
<tr>
<td>MDQ</td>
<td>.90</td>
<td>DX</td>
<td>No</td>
<td>No</td>
<td>5 min</td>
<td>No</td>
</tr>
<tr>
<td>GAD-7</td>
<td>.90</td>
<td>DX</td>
<td>No</td>
<td>No</td>
<td>5 min</td>
<td>No</td>
</tr>
<tr>
<td>BAI</td>
<td>.92</td>
<td>DX, PSY</td>
<td>Yes</td>
<td>No</td>
<td>5–10 min</td>
<td>Yes</td>
</tr>
<tr>
<td>Substance use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUDIT</td>
<td>.86</td>
<td>PSY, DX</td>
<td>No</td>
<td>No</td>
<td>5 min</td>
<td>No</td>
</tr>
<tr>
<td>MAST</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>10–15 min</td>
<td>No</td>
</tr>
<tr>
<td>SASSI</td>
<td>.93</td>
<td>DX, PSY</td>
<td>No</td>
<td>No</td>
<td>20 min</td>
<td>Yes</td>
</tr>
<tr>
<td>Eating pathology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDE-Q</td>
<td>.72–.95</td>
<td>DX, PSY, EB</td>
<td>Yes</td>
<td>No</td>
<td>5–10 min</td>
<td>No</td>
</tr>
<tr>
<td>QESP</td>
<td>NA</td>
<td>DX, PSY, EB</td>
<td>Yes</td>
<td>No</td>
<td>5 min</td>
<td>No</td>
</tr>
<tr>
<td>TFEQ</td>
<td>.73–.87</td>
<td>DX, EB</td>
<td>Yes</td>
<td>No</td>
<td>20 min</td>
<td>No</td>
</tr>
<tr>
<td>BES</td>
<td>.90</td>
<td>DX, PSY, EB</td>
<td>Yes</td>
<td>No</td>
<td>5 min</td>
<td>Yes</td>
</tr>
<tr>
<td>EDE-3</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>Yes</td>
<td>20 min</td>
<td>Yes</td>
</tr>
<tr>
<td>NEQ</td>
<td>.70</td>
<td>EB</td>
<td>Yes</td>
<td>No</td>
<td>5–10 min</td>
<td>No</td>
</tr>
</tbody>
</table>

Note. WL = weight loss; ADH = adherence; PSY = other measures of psychopathology; EB = eating behaviors; DX = diagnoses.
of validity scales, which provide information on the extent that patients are responding consistently to items of the test, overreporting symptoms, or underreporting symptoms. Disadvantages to using broadband assessment instruments typically include the cost per administration, patient burden, and lack setting specific content area, such as eating behaviors. This section aims to briefly review both the content and psychometric properties of broadband assessment instruments used in bariatric surgery settings.

**MMPI-2 (MMPI-2).** The MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; Butcher et al., 2001) is the most widely used broadband assessment instrument in bariatric surgery clinics (Bauchowitz et al., 2005; Fabricatore et al., 2006; Walfish et al., 2007). The MMPI-2 consists of 587 true/false items that can be administered in 60–90 min. The test contains nine validity scales to assess protocol validity and 10 clinical scales that assess broadband psychological constructs. The Extended Score Report and Minnesota Report also scores items on nine Restructured Clinical Scales, 31 Harris-Lingos Scales, 15 Content Scales, 27 Content Component Scales, 15 Supplemental Scales, and five Psychopathy—Pathology–5 Scales. The MMPI-2 also has published normative data available for bariatric surgery patients and is included in the Manual for Administration and Scoring. A Spanish translation is also available.

When used in bariatric surgery settings, the test demonstrates adequate to good reliability coefficients with the exception of Clinical Scale 6 (Paranoia; Wygant et al., 2007), though this result is likely because of a low base rate of thought disorders in these samples (Mitchell, Selzer, et al., 2012). Walfish (2007) demonstrated that elevated scores on the underreporting scales of the MMPI-2 not only suppress scores on the clinical scales of the test, but also suppress scores on other self-report inventories used in the test battery. Therefore, if a patient underreports on the MMPI-2, it is supported that they also underreport across the psychological evaluation. This finding is also supported in other samples (Forbey, Lee, Ben-Porath, Arbisi, & Gartland, 2013). In terms of external validity, the MMPI-2 clinical scales yield good convergent validity and adequate discriminant validity with presurgical criteria gathered from the clinical interview (Wygant et al., 2007). In relation to predictive validity, empirical evidence suggests that high scores on Clinical Scales 3 (hysteria), 5 (masculinity/femininity), and 6 (paranoia) in addition to the Health Concerns Content Scale (HEA), and the F (infrequency) Validity Scale differentiated between patients who lose <50% of their excess weight 1 year after a Roux-en-Y Gastric Bypass. Similar to studies using structured clinical interviews (de Zwaan et al., 2011), Maddi et al. (2001) reported a reduction in MMPI-2 clinical scale elevations from presurgery to up to 1-year postsurgery, suggesting improvements in psychosocial functioning after bariatric surgery.

**MMPI-2 Restructured Form (MMPI-2-RF).** The MMPI-2-RF (Ben-Porath & Tellegen, 2008/2011; Tellegen & Ben-Porath, 2008/2011) is the newest version of MMPI instruments. The MMPI-2-RF contains 338 items (true/false format) taken from the MMPI-2 and yields a 35–50 min administration time. The rationale for restructuring the MMPI-2 was to improve on the psychometric properties of the MMPI-2, reduce patient burden, and offer a straightforward interpretative strategy. Moreover, the items are scored on nine validity scales and 42 substantive scales. The instrument is designed in congruence with current, theoretical models of psychopathology (Kotov et al., 2011; Krueger & Tackett, 2006; Sellbom, Ben-Porath, & Bagby, 2008). Specifically, the test assesses emotional, thought, and behavioral functioning in addition to somatic/cognitive and interpersonal problems. Each substantive scale offers diagnostic considerations and treatment implications/recommendations. The MMPI-2-RF contains normative data for bariatric surgery patients that has been replicated (Marek et al., 2013; Tarescavage, Wygant et al., 2013). A disadvantage is that a Spanish translation is not yet available for the MMPI-2-RF; however, the Spanish translation of the MMPI-2 can be administered and scored as a MMPI-2-RF.

Although usage statistics on how many bariatric surgery programs use the MMPI-2-RF are not yet available, the scale scores of the MMPI-2-RF have demonstrated better psychometric properties when compared with the scale scores of the MMPI-2 when used in bariatric surgery settings (Marek et al., 2013; Tarescavage, Wygant et al., 2013; Wygant et al., 2007). Specifically, the scales of the instrument yield better reliability coefficients and better discriminant validity when used in bariatric settings as compared with the scales of the MMPI-2 (Tarescavage, Wygant et al., 2013; Wygant et al., 2007). In addition, the scales of the instrument demonstrate good convergence and incremental validity with diagnoses derived from a semistructured interview (Marek, Ben-Porath et al., 2014a; Marek, Ben-Porath, Ashton, & Heinberg, 2014b; Marek et al., 2013). Using a stepdown, hierarchical regression procedure with external criteria, Marek, Ben-Porath, Sellbom, McNulty, and Heinberg (2015) demonstrated that the test scores are valid regardless of gender, age, or ethnicity. Moreover, Tarescavage, Windover et al. (2013) provide interpretation strategies for the suicidal/death ideation scale and the substance abuse scale for identifying bariatric surgery candidates who are at an elevated risk of suicidal ideation and substance use, respectively.

In terms of predictive validity, recent research (Marek, Ben-Porath, Merrell, Ashton, & Heinberg, 2014) supports that presurgical MMPI-2RF scores predict early adjustment related difficulties as soon as 1 and 3 months after bariatric surgery in the areas of somatization, distress, and maladaptive eating patterns. In a longer-term outcome study, presurgical scores on the behavioral/externalizing dysfunction scales incrementally predicted suboptimal 1-year weight loss outcomes and appointment adherence after controlling for presurgical BMI and demographic variables in patients who underwent a Roux-en-Y procedure (Marek et al., 2015).

**Personality Assessment Inventory (PAI).** The PAI (Morey, 1991; Morey, 1996, 2007) is a self-report inventory containing 344-items. Items are Likert-scale format and produce 22 nonoverlapping scales consisting of four validity scales, 11 clinical scales, five treatment scales, and two interpersonal scales. An additional 10 subscales are also available to aid in interpretation. The clinical scales assess psychopathology in three spectrums: neurotic, psychotic, and behavioral/impulse control. The treatment scales assess the degree to which a patient is at higher risk for self-harm or harm to others, how patients respond to environmental factors, and motivation for treatment. The interpersonal scales assess the degree to which a patient is warm/affiliative versus cold/rejecting and dominating/controlling versus submissive. Test administration is approximately 50–60 min, and a Spanish version of the test is available. Normative data for use with bariatric surgery patients have been published (Corsica, Azarbad, McGill, Wool, & Hood,
patients (Ransom, Ashton, Windover, & Heinberg, 2010). The measure assesses global psychological distress. The items are Likert-scale format such that higher scores reflect greater severity. Patients are presented with a list of symptoms and asked to report how often each symptom has affected them in the past 7 days. Higher scaled scores are indicative of greater symptom intensity on nine different subscales including areas such as anxiety, depression, and somatization. The measure also details three global scales to assess response biases is also a serious limitation to the use of the SCL-90–R, as patients seeking surgery may be motivated to present themselves favorably in the evaluation setting (Ambwani et al., 2013).

**Millon Behavioral Medicine Diagnostic (MBMD).** The Millon Behavioral Medicine Diagnostic (MBMD; Millon & Antoni, 2006) is a broadband instrument designed to be used in behavioral health settings. The test contains 165 true/false items that are scored on three validity scales and 32 clinical scales that are grouped into the following domains: response patterns, negative health habits, psychiatric indications, coping styles, and stress moderators. The test materials include bariatric surgery normative data and have an interpretative report specific for use with bariatric surgery patients (Millon, Antoni, Millon, Minor, & Grossman, 2007). The test can be administered in 20–25 min, has online scoring available as an option, is associated with a fee, and currently offers a Spanish translation.

Although normative data and a comprehensive clinical report are available for bariatric surgery candidates, limited psychometric data are published. Walfish, Wise, and Streiner (2008) reported that at least half of the MBMD scales yielded internal consistency estimates less than .70 and argued that the validity of scales scores of the MBMD in the bariatric surgery population is, therefore, limited. As reviewed earlier, reliability sets the upper bounds for validity. To this end, it is possible that some scales on the MBMD will not produce good validity coefficients in this setting owing to poor internal consistency (see Walfish, Wise, & Streiner, 2008 for a critical examination of these limitations of the MBMD).

**Millon Clinical Multiaxial Inventory—III (MCMI—III).** The MCM-I-III (Millon, Millon, Davis, & Grossman, 2009) is the latest iteration of MCMI instruments. The instrument is composed of 175 true/false questions scored on five validity scales and 24 personality and clinical scales that purport to assess DSM–IV–TR diagnoses. The test has demonstrated adequate to good reliability and validity coefficients and contains normative data for psychiatric inpatients and patients seeking mental health treatment. A Spanish translation is available, and the test can be administered in 25–30 min. Moreover, a new version of the test, the MCMI-IV, is currently in development.

Although the test manual contains psychometric information for use with psychiatric inpatients and patients seeking mental health treatment, the research supporting the instruments psychometric properties for use with bariatric surgery patients is limited. To date, no reliability or normative data supporting the use of the MCMI-III with bariatric surgery candidates have been published. In terms of external correlates, Belanger, Wechsler, Nademin, and Virden (2010) reported associations between the scales of the MCMI-III and short-term weight loss outcomes in a sample of 143 bariatric surgery patients. The researchers reported scales schizoid, schizotypal, and compulsive predicted less weight loss at 6-months postsurgery. Martínez, González, Vicente, van der Hofstadt Román, and Rodríguez-Marín (2013) cluster analyzed MCMI-III scale scores from a sample of 53 bariatric surgery patients and correlated the clusters with external criteria, such as measures of depression and anxiety, providing some construct validity data for the use of the measure with bariatric patients.
Basic Personality Inventory (BPI). The Basic Personality Inventory (BPI; Jackson, 1976) is a broadband instrument that contains two validity scales and 12 clinical scales scored from 240 true/false items. The instrument can be administered in approximately 35 min and online scoring is available. Non-English translations are unavailable. The test has limited psychometric support with the bariatric surgery population. Lanyon and Maxwell (2007) reported associations between presurgical BPI scores and weight loss 1 year after bariatric surgery. Low alienation scores were associated with better weight loss outcomes. No data on the instruments psychometrics (such as reliability coefficients, normative data, or other reports of validity) have been reported.

Narrowband Instruments

Many of the more broadband measures described above are limited by a lack of assessment of disordered eating. Moreover, all of the above instruments yield a longer administration time compared with narrowband measures and are associated with a fee. In situations when administering a broadband instrument is not feasible or fails to assess a relevant domain (i.e., eating disorder behaviors), narrowband instruments can be a good substitute. In addition, narrowband measures are also suitable for assessing change in symptoms across time and can be administered quickly. However, it is not recommended to comprise an assessment battery solely on the narrow-band measures described below because none of the instruments discussed contain validity scales and are missing content that a broadband measure can capture, such as somatization, interpersonal functioning, or thought disorder. Should clinicians begin to use narrowband measures that assess for somatization, interpersonal functioning, or thought disorder in bariatric surgery settings, it is recommended that reliability, validity, and normative data should be published and support the measures’ utility prior to implementing them in practice.

Depression/anxiety measures.

Beck Depression Inventory-II (BDI-II). The Beck Depression Inventory-II (BDI-II; Beck et al., 1996) is a 21-item self-report scale of depressive symptoms commonly used in bariatric surgery evaluations (Bauchowitz et al., 2005). The measure assesses severity of depression, with scores of 0–13 suggesting no depression, 14–19 mild depression, 20–28 moderate depression, and ≥29 severe depression. Research suggests that the BDI-II is an adequate screening measure for depression in bariatric populations (Hayden, Brown, Brennan, & O’Brien, 2012; Wadden et al., 2006); however, the BDI-II score tends to assess more affective symptoms of depression and general somatic complaints that are independent of what is typically observed in patients with major depressive disorder (de Zwaan et al., 2011). Advantages of the BDI-II include ease of administration/scoring; disadvantages include cost and lack of discriminant validity when used in bariatric surgery settings. Although it has been recommended that a cutoff score of 13 for further depression screening in bariatric candidates (Hayden et al., 2012), discriminant properties of the BDI-II scores when used in this setting warrant further research. For example, Krukowski, Friedman, and Applegate (2010) suggest that the BDI-II also assesses somatic complaints (e.g., sexual dysfunction, chronic pain) in bariatric settings and the results of a somatic factor have been partially replicated in different bariatric surgery samples (Hall et al., 2013; Udo, McKee, & Grilo, 2015). Therefore, the instrument’s discriminant validity should be further assessed.

The Patient Health Questionnaire-9 (PHQ-9). The Patient Health Questionnaire-9 (PHQ-9; Kroenke & Spitzer, 2002) is a nine-item screening tool for depression that has been used extensively in medical populations, including bariatric patients. The instrument has demonstrated good convergent validity with DSM diagnoses in other populations, yields a short administration and scoring time, and carries no cost. PHQ-9 scores have demonstrated good reliability, validity, and utility as a screening tool for depression when used with bariatric surgery patients (Cassin et al., 2013) and demonstrated good concurrent validity when compared to the BDI-II in bariatric surgery settings (Schutt, Kung, Clark, Koball, & Grothe, 2015). Researchers suggest a cutoff score of 15 for further screening depression in bariatric candidates noting the frequency of somatic complaints in this population (Cassin et al., 2013).

The Center for Epidemiologic Studies Depression Scale (CES-D). The Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) is a 20-item scale for depression. The scale uses self-statements (e.g., “I felt hopeful about the future”) and respondents rate how frequently each item applied to them over the course of the past week. Ratings are based on a 4-point Likert scale ranging from 0 = rarely or none of the time (less than 1 day) to 3 = most or all of the time (5–7 days). The scale score demonstrates good reliability across populations, and has the advantage of being publically available. Disadvantages include no studies assessing the psychometric properties of the CES-D scale score in bariatric settings and the inclusion of items pertaining to somatic symptoms common in medical populations such as bariatric surgery candidates may impair the instruments discriminant properties.

Mood Disorders Questionnaire (MDQ). Bipolar symptoms are as prevalent in bariatric surgery patients as compared to the general population (Kalarchian et al., 2007) and may warrant further screening. The Mood Disorders Questionnaire (MDQ; Hirschfeld et al., 2000) is a bipolar screening tool consisting of 13 items that has been studied in bariatric populations (Grothe et al., 2014). A cutoff score of <7 yields good sensitivity in detecting bipolar spectrum-related disorder in psychiatric outpatient samples (Hirschfeld et al., 2000). The MDQ scale score demonstrates good reliability coefficients and yields good convergent validity coefficients with bipolar diagnoses.

Generalized Anxiety Disorder-7 (GAD-7). The GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006) is a seven-item measure designed to screen for anxiety. Its scale score has good reliability, validity, and is closely tied to DSM diagnostic criteria for generalized anxiety disorder. The instrument yields good reliability coefficients when used in bariatric surgery settings (de Zwaan et al., 2014). There is limited research with the instrument in bariatric settings. One study reported patients who undergo body contouring surgery after having bariatric surgery report a reduction in GAD-7 scores (Azin et al., 2014). A second study by de Zwaan et al. (2014) reported no significant differences on the GAD-7 between postbariatric surgery patients who reported dissatisfied body image and had an abdominoplasty versus patients who did not undergo an abdominoplasty. Additional psychometric evaluations of validity within this population have yet to be published.
The Beck Anxiety Inventory (BAI). The Beck Anxiety Inventory (BAI; Beck & Steer, 1993) is a 21-item, multiple-choice, self-report inventory that measures the severity of anxiety. The items consist of four statements that range in severity for each anxiety symptom. Values for each item are summed yielding an overall or total score for all 21 symptoms ranging from 0 to 63 points. A total score of 0–7 is interpreted as a minimal level of anxiety; 8–15 as mild; 16–25 as moderate; and 26–63 as severe. The BAI score has demonstrated good reliability and convergent validity when used with bariatric surgery patients (Edwards-Hampton et al., 2014).

Substance use measures.

Alcohol Use Disorders Identification Test (AUDIT). The Alcohol Use Disorders Identification Test (AUDIT; Babor, Higgins-Biddle, Saunders, & Monteiro, 2001; Saunders et al., 1993) was developed by the World Health Organization to screen individuals for harmful or hazardous drinking patterns. It is a 10-item oral interview or self-report questionnaire that measures hazardous alcohol use (three items; e.g., dependence symptoms (three items), and harmful alcohol use (four items) in the past 12 months. A total score is calculated by summing each item with a range from 0–40, with higher scores reflecting more harmful alcohol use. A score of 8 or higher indicates harmful or hazardous drinking and a possible alcohol use disorder. The AUDIT score has shown good reliability and validity. Advantages include brief administration (5 min), easy scoring, and free of charge as it is publicly available. The AUDIT has been used to screen for alcohol use disorders in bariatric surgery populations (King et al., 2012; Mitchell, Steffen, et al., 2014).

Derived from the AUDIT, the AUDIT-C (Alcohol Use Disorders Identification Test—Consumption; Bush, Kivlahan, McDonell, Fihn, & Bradley, 1998) is a brief, modified version of the AUDIT that assesses alcohol consumption in just three items. Although the instrument is short, one study reported good sensitivity and specificity in differentiating bariatric surgery patients who met diagnostic criteria for an alcohol use disorder as compared with those who did not (Suzuki, Haimovici, & Chang, 2012). The AUDIT-C scale score requires additional psychometric research to further attest to its validity.

The Michigan Alcoholism Screening Test (MAST). The Michigan Alcoholism Screening Test (MAST) was designed to measure alcohol use and related problematic behaviors (Selzer, 1971). It is a 25-true/false item, self-report inventory in which each item is scored 0, 1, 2 or 5 points. Two items score 2 additional points for each prior arrest. A total score is calculated by summing all item scores with a range of 0–53. Scores ranging from 0–3 indicate no problem, 4 indicate an early or middle problem drinker, and 5 or greater indicate a problem drinker (alcoholic). The scale score of the MAST has demonstrated good reliability and validity when used in clinical settings (Storgaard et al., 1994; Zung, 1979). An advantage to this measure is that it is free of charge as publically available. Disadvantages include the longer length of time to administer and more complicated scoring compared to other screening instruments available. Additionally, the MAST inquires about any history of problems (e.g., questions include “ever”) and may be measuring lifetime problematic alcohol use without adequately differentiating from current use (Storgaard et al., 1994). Last, although the MAST has been used in clinical practice, no studies to date have evaluated the psychometric properties when used in bariatric surgery settings.

Substance Abuse Subtle Screening Inventory-3 (SASSI-3). The Substance Abuse Subtle Screening Inventory-3 (SASSI-3; Miller, Roberts, Brooks, & Lazowski, 1997) was designed to identify those at risk for a substance use disorder. The test purports to use a subtle approach to assessing substance use in that items inquiring directly about substance use are limited (Clements, 2002). The SASSI-3 is a 93-item, self-report inventory with 67 true/false items that compose eight subscales: symptoms of substance misuse, obvious attributes, subtle attributes, defensiveness, supplemental addiction measure, family versus control subjects, correctional, and random answering pattern. There are also 26-items that more directly inquire about substance use with response options of never, once or twice, several times, or repeatedly (Miller et al., 1997). Scale scores of the SASSI-3 have demonstrated good reliability and validity (Lazowski, Miller, Boye, & Miller, 1998). A potential advantage to this measure is the ability to examine subscales and a potential validity subscale of random responding. Disadvantages include the cost of administration, longer administration time, and more complicated scoring compared with other screening instruments available. Another disadvantage is the limited research on the psychometric properties of this scale and discrepant findings regarding sensitivity (Clements, 2002). For example, Clements (2002) found a low sensitivity level (.65) for identifying alcohol dependence in college students, although improved when only examining the face valid items of the SASSI-3 (.89). Last, although the SASSI-3 has been used in clinical practice, no studies to date have evaluated the psychometric properties when used in bariatric surgery settings.

Eating behavior measures. Noting the prevalence of disordered eating behavior in this population (Colles et al., 2008; Marek et al., 2013; Mitchell, Selzer, et al., 2012) and evidence that the persistence of eating disorders psychopathology (e.g., loss of control eating, grazing) may contribute to less favorable weight loss outcomes, assessment of these behaviors should be included in any presurgical psychological evaluation. Historically, well-validated instruments were developed and normed for patients with lower BMI’s or for those presenting with anorexia nervosa or bulimia nervosa. However, in recent years, a number of self-report questionnaires initially developed for and validated in eating disordered populations have been utilized to assess eating disordered symptoms in obese populations or in obese patients presenting for weight loss surgery.

A recent systematic review of the literature on eating disordered assessment in bariatric surgery candidates identified over 34 questionnaires in published use (Parker, Mitchell, O’Brien, & Brennan, 2015). The most commonly reported and used in clinical practice are described below. Although many of the following eating disorder screening instruments have reported psychometrics in a weight loss surgery population, it is essential to confirm eating disorder diagnoses with a clinical interview or structured research based assessments such as the Eating Disorder Examination (EDE; Fairburn, 2008) or the Structured Clinical Interview for DSM (SCID; First, Williams, Karg, & Spitzer, 2015) as numerous studies have demonstrated highly divergent prevalence depending upon the instrument used and relatively poor convergence with research based clinical interviews (Mitchell et al., 2010; Parker et al., 2015).
Eating Disorders Examination Questionnaire (EDE-Q). The Eating Disorders Examination Questionnaire (EDE-Q; Fairburn, 2008) is the 32-item self-report version of the interviewer-based EDE with subscales measuring restraint, eating concern, weight concern, and shape concern. The measure is publically available and can be administered in approximately 5–10 min. This commonly utilized instrument has been validated in bariatric surgery patients and its scale scores have demonstrated good internal consistency (.72–.95) and convergent validity despite research supporting different factor structures (Hrabosky et al., 2008). There are strong correlations between the scales of the EDE-Q and the interviewer-based scores of the EDE in bariatric samples (Kalarchian, Wilson, Brolin, & Bradley, 2000). All subscales (except restraint) have moderate to strong correlations with measures of body image, depression, and self-esteem when evaluated using bariatric samples (Elder et al., 2006).

Questionnaire of Eating and Weight Patterns–Revised (QEWP-R). The Questionnaire of Eating And Weight Patterns–Revised (QEWP-R; Spitzer, Yanovski, & Marcus, 1993) is a 28-item measure assessing a variety of behavioral aspects of disordered eating including objective binge eating episodes, weight history, and body image. This is the most commonly utilized measure within the literature. The scores have shown good convergence (κ = .57) with the SCID for diagnosing BED (de Zwaan et al., 1993) in obese treatment-seeking women. However, its convergence with binge eating status on the EDE-Q in bariatric patients was far lower (κ = .26) and the EDE-Q was better able to differentiate between individuals who binged one time per week and those who reported no or infrequent binge eating. The EDE-Q was also more sensitive in identifying overconcern with body weight and shape (Elder et al., 2006).

The Three-Factor Eating Questionnaire (TFEQ). The Three-Factor Eating Questionnaire (TFEQ; Stunkard & Messick, 1985) is a 51-item instrument designed to assess three factors of interest: restraint, hunger, and disinhibition. However, confirmatory factor analytic studies have not consistently supported these three factors. It is also frequently cited in the bariatric literature (Parker et al., 2015; Sarwer et al., 2012; Sarwer et al., 2008). In an obese population, the three purported factors of the TFEQ demonstrate adequate internal consistency—Cronbach’s α = .87 (restraint), .73 (hunger), and .80 (disinhibition)—and convergence with disordered eating behavior and discriminant validity (Allison et al., 1992). Scores on the TFEQ have been previously reported in bariatric patients and it distinguished between binge and nonbinge eaters (de Zwaan et al., 2003; Kalarchian, Wilson, Brolin, & Bradley, 1998).

Binge Eating Scale (BES). Another commonly utilized instrument in bariatric populations is the Binge Eating Scale (BES; Gormally, Black, Daston, & Rardin, 1982) which consists of 16 multiple-choice items designed to assess binge eating severity. The scale of the measure has been found to have construct validity in its ability to distinguish between minimal, moderate, and severe binge eating problems. Clinical cut-offs of binge eating severity have been established for the BES (Grupski et al., 2013) and internal consistency has been reported in bariatric populations (Cronbach’s α = .90; Ashton, Drerup, Windover, & Heinberg, 2009). A two-factor structure (emotions/cognitions and behavioral manifestations) has been supported in bariatric samples in two different investigations (Hood, Grupski, Hall, Ivan, & Corsica, 2013; Marek, Tarescavage, Ben-Porath, Ashton, & Heinberg, 2015). However, it is recommended to be used unidimensionally because both factors are highly correlated (r = .89) and the behavioral manifestation factor lacks incremental validity (Marek, Tarescavage, Ben-Porath, Ashton, & Heinberg, 2015).

The Eating Disorders Inventory–III (EDI-III). The Eating Disorders Inventory–III (EDI-III; Garner, 2004) is a 91-item Likert-itemed scale with 11 subscales measuring drive for thinness, bulimia, body dissatisfaction, ineffectiveness, perfectionism, interpersonal distrust, introceptive awareness, maturity fears, asceticism, impulse regulation, and social insecurity. It has been validated in obese populations but not among bariatric surgery candidates. In spite of this, a number of studies have utilized earlier versions of this instrument or specific subscales of EDI instruments (cf., Adami, Meneghelli, Bressani, & Scopinaro, 1999; Thonney, Pataky, Badel, Bobbioni-Harsch, & Golay, 2010).

Night Eating Questionnaire (NEQ). The Night Eating Questionnaire (NEQ; Allison et al., 2008) is a 14-item, 5-point Likert scale used to measure the severity of symptoms associated with night eating syndrome. It measures four factors associated with night eating including nocturnal ingestion, evening hyperphagia, morning anorexic, and mood/sleep. It has adequate internal consistency (α = .70), convergent validity, and has been validated in weight loss surgery candidates (Allison et al., 2008).

Eating measures for further study. More recently, two measures have been described that more specifically measure loss of control. The Loss of Control Eating Scale (LOCES; Latner, Mond, Kelly, Haynes, & Hay, 2014) is a 24-item Likert-scaled instrument designed to assess the loss of control components of binge eating. The construct of “loss of control” has been more predictive of psychopathology and distress than the amount of food consumed. The scales of the LOCES have very good internal consistency in a nonclinical sample (Cronbach’s α = .96), converges well with other measures of disordered eating and psychopathology and has three factors: (a) behavioral aspects, (b) cognitive/dissociative aspects, and (c) positive euphoric aspects of loss of control eating. A seven-item LOCES-Brief is also available with similar psychometric properties.

Similarly, the Eating Loss of Control Scale (ELOCs) was published the same year (Blomquist et al., 2014). This single-factor 18-item Likert-scaled instrument demonstrates good internal consistency (Cronbach’s α = .90) and good convergence with measures of eating and psychopathology. The scale score was normed on university students whereas the ELOCs was normed on obese persons with binge eating disorder. Specific use in bariatric patients have not been reported, but further research on these measures should be considered in this population.

Although structured interviews remain ideal for diagnosing eating disorders in this population, the previously described screening instruments may be a helpful adjunct to more broad-based measures of personality and psychopathology. In their systematic review, Parker, Mitchell, O’Brien, and Brennan (2015) recommend that the BES be used with caution given tendency to over-diagnose but its strength is that it was developed specifically for obese individuals and that it can be a helpful screening instrument. They also recommend the NEQ for screening night eating behaviors although more psychometric development is needed. Finally, the EDE-Q subscales were recommended as beneficial in assessing...
the severity of eating pathology although, like the BES, its scale scores may be higher than the interview based EDE.

Discussion

Despite the AACE/ASMB/S/TOS Guidelines for the Clinical Management of Bariatric Surgery in addition to guidelines set out by NIH in 1991 recommending that assessment of bariatric surgery candidates include some form of objective psychological testing (Mechanick et al., 2008, 2013; NIH, 1991), survey data suggest only about two thirds of bariatric surgery clinics adhere to including a psychological measure in their evaluation. As reviewed, some instruments being utilized, such as the MBMD and the MAST, have limited to no psychometric data available to justify their use in bariatric surgery settings. The previous published guide on using objective testing in bariatric surgery clinics (LeMont et al., 2004) only provided a description of instruments available for use and did not include a review of their psychometric properties.

Psychological testing is an important component of the psychological evaluation, adding normative data and additional lines of evidence to provide support for diagnoses and treatment planning/recommendations. Psychological testing can also provide information about the patients’ response style, such as underreporting, that can help to provide a clearer picture of the patient. Providers of presurgical psychological evaluations are strongly recommended to include psychological testing in addition to the clinical interview to provide a more robust psychological evaluation.

Because of the broad number of psychosocial domains to assess in bariatric candidates, including a psychometrically reliable and valid broadband instrument, such as the MMPI-2-RF or PAI, is recommended. Although both of these instruments are associated with a cost and do not assess eating pathology, they adequately assess most domains of psychosocial functioning relevant to the assessment of bariatric surgery candidates. Moreover, they can be administered in under an hour, have validity scales to detect underreporting, and scoring software is available. Both instruments also have bariatric normative data. To date, there is ample literature suggesting the MMPI-2-RF is suitable to be included as a broadband instrument in bariatric surgery evaluations. The test has demonstrated replicable normative data, good reliability and validity coefficients, is associated with eating behaviors (though does not directly assess it), has predictive utility, and is not demographically biased (Marek, Ben-Porath et al., 2014b; Marek, Ben-Porath et al., 2015; Marek et al., 2013; Marek et al., 2015; Tarescavage, Wygant et al., 2013). It is important to note that although the scale scores of the MMPI-2-RF have demonstrated good reliability, validity, and practical utility, associations with longer term weight loss (i.e., 3 years, 5 years, 10 years, etc.) and other external criteria at these postoperative follow-ups are nonexistent at this time. Although literature on the PAI is not as extensive compared with the literature on the MMPI-2-RF, literature does support the scale scores of the PAI are reliable, valid, and have normative data (Ambwani et al., 2013; Corsica et al., 2010). Moreover, the results associated with elevations on the Positive Impression Management scale (an underreporting validity scale) are similar to those reported for the MMPI-2-RF underreporting scales (Ambwani et al., 2013; Marek et al., 2015; Tarescavage, Windover et al., 2013). However, it is important for clinicians and researchers to publish additional data to support the use of the PAI in this setting.

Another strength of using some of the broadband instruments reviewed is the ability to assess protocol validity (i.e., the extent to which individuals engage in random or fixed responding, overreporting of symptoms/behaviors, or underreporting of symptoms/behaviors). A proportion (~30%) of bariatric surgery patients minimize psychopathology at their preoperative evaluation, most notably impulse-control and sensation-seeking problems (Ambwani et al., 2013; Marek, Tarescavage et al., 2015; Tarescavage, Windover et al., 2013; Walfish, 2007). Moreover, it can be inferred that if an individual over- or underreports their psychological functioning on instruments such as the MMPI-2-RF, they also engage in a similar response style across other assessment measures (Forbey, Lee, Ben-Porath, Arbisi, & Garland, 2013). Inclusion of measures with embedded with validity indices can aid in interpretation of test results and decision making.

Currently, the MBMD or the MCMII, due to inadequate literature supporting their psychometric properties when used in the bariatric settings, are considered secondary choices. Although not recommended for clinical practice at this juncture, additional research to support their use is recommended. It is recommended that clinicians using the MBMD or MCMII in practice publish studies further examining the psychometric properties of these instruments in bariatric surgery settings to better support their utility. The MMPI-2, though widely used in this setting, is also being considered a secondary choice due to scale scores yielding poorer reliability and discriminant validity coefficients as compared to the newer MMPI-2-RF. The SCL–90–R has evidence for reliability and bariatric norms; however, limitations include the lack of validity scales. Research outside of bariatric surgery suggests the test primarily measures psychological distress as opposed to the specific symptom categories that is purports to assess (Cyr, McKenna-Foley, & Peacock, 1985; Hayes, 1997; Vassend & Skrondal, 1999).

Of the mood/anxiety narrowband measures, the PHQ-9 is a strong choice for depression. The PHQ-9 yields good psychometric properties when used in bariatric surgery settings, is brief and has no cost (Cassin et al., 2013). The PHQ-9 is also suitable for an assessment of change across time for the aforementioned reasons. The BDI-II is widely used in this setting, but its discriminant validity in bariatric surgery settings warrants further study. The BAI is a good choice to measure anxiety, though it carries the same limitations and considerations mentioned for the BDI-II. The GAD-7, the CES-D, and the MDQ warrant further research to established psychometric properties for use in bariatric surgery settings.

In terms of substance use measures, the AUDIT and AUDIT-C have been psychometrically established in bariatric surgery settings and are commonly used (King et al., 2012; Mitchell, Steffen, et al., 2014; Suzuki et al., 2012). The MAST and SASSI-3 warrant further research before a recommendation for clinical use with bariatric patients can be made.

Clinicians should also include an eating behavior measure in their battery. Many of the reviewed narrowband eating measures have a short administration time. The EDE-Q is recommended as an eating related self-report measure. The EDE-Q is multifaceted, publically available, yields a short administration time, and demonstrates adequate to good reliability coefficients, is associated
with diagnoses derived from structured clinical interviews, and has been associated with other related external criteria, such as body image (Elder et al., 2006; Hrabosky et al., 2008; Kalarchian et al., 2000). Though not multifaceted, the BES also demonstrates strong reliability coefficients, good construct validity, and is associated with binge eating severity (Ashton et al., 2009; Grupski et al., 2013; Hood et al., 2013; Marek, Tarescavage, Ben-Porath, Ashton, & Heinberg, 2015). A disadvantage of the BES is a narrow focus on binge eating severity. The NEQ is recommended for patients reporting poor eating behaviors associated with night eating syndrome. The NEQ score has been validated for use in bariatric surgery settings (Allison et al., 2008) and can be used in addition to a broader assessment of eating behaviors. Although the scores of the TFEQ have demonstrated adequate reliability, convergent, and discriminant validity in bariatric surgery settings (Allison, Kalinsky, & Gorman, 1992), its factorial validity has not been able to replicate (Karlsson, Persson, Sjostrom, & Sullivan, 2000). Psychometric information for the QEW-P-R are generally poor and the EDE-Q provided incremental validity over the QEW-P-R (de Zwaan et al., 1993; Elder et al., 2006).

In considering psychological testing batteries for weight loss surgery evaluations, clinicians are encouraged to consider the psychometric properties, availability of bariatric norms, validity of domain assessment, and practical considerations such as patient burden, cost, and ease of administration. A psychometric summary of all measures discussed are located in Table 1. Different practices may have different time and practical constraints, making selection of a testing battery differ across settings. In addition, clinicians may choose additional psychological tests based on particular patient characteristics, multidisciplinary team questions/ referrals, or clinical judgment.

Overall, the inclusion of a broadband instrument and narrow-band measures of eating, mood, and substance use are suggested as the foundation of an assessment battery in bariatric surgery evaluations. Of those reviewed, the MMPI-2-RF and the PAI are broadband measures recommended for practice due to psychometric support of those instruments in bariatric surgery settings. These instruments can be administered in an hour, have literature supporting the psychometric properties of their scale scores, offer validity scales, and have published normative data. The EDE-Q, BES, or NEQ are recommended to assess for various eating-related pathology and behaviors. Additional measures such as the PHQ-9 and the AUDIT may augment the battery and provide information on depression state and alcohol use. They are also brief to administer if change across time is of interest. In summary, use of psychological assessment instruments that have been validated in bariatric surgery settings enrich the psychological evaluation and provide additional avenues to help patients achieve successful surgical results.

References
BARIATIC SURGERY ASSESSMENT RECOMMENDATIONS


Received September 8, 2015
Revision received December 16, 2015
Accepted December 23, 2015