CHAPTER 4

Assessment of Body Image

Hemal P. Shroff, Rachel M. Calogero, and J. Kevin Thompson

The field of body image has experienced a phenomenal growth in recent years. In 2004, a new journal, *Body Image: An International Journal of Research,* was launched by Elsevier to provide a forum for publishing papers in this area. The field of body image was once closely linked to that of eating disorders, but the construct is now often studied in a variety of fields, including, oncology, dentistry, surgery, sociology, nutrition, and obesity (Cash & Pruzinsky, 2002). In addition, measurement of body image in younger samples and men has received a great deal of attention in recent years (Cafri & Thompson, 2007; Yanover & Thompson, 2009). The expansion of the field in general has been paralleled by a dramatic increase in the number of measures developed to assess some dimension of "body image."

In this chapter, we aim to briefly cover the broad categories of measures, discuss recent innovations, and outline important methodological issues. It is not possible for us to discuss all the measures in detail; therefore, we will focus our discussion on those measures that have been widely used and received rather extensive psychometric evaluation. We provide information regarding many of the available measures in Table 4.1. Within the text and in the table, we categorize measures using standard terminology into the following dimensions of body image: subjective and affective, cognitive, and behavioral. We also include a subsection on new and widely used measures that examine sociocultural and interpersonal influences on body image because of their relevance for assessing associated features of body image disturbance that may have treatment indications. It should be noted that we do not discuss perceptual measures in this chapter because they are seldom used in clinical practice, and there is an ongoing debate regarding whether these measures provide a distinct index of perceptual body image (J. K. Thompson & Gardner, 2002). Also, at the end of this chapter, we offer some detailed guidelines for choosing a measure based on gender, age, and ethnicity.

It is important to keep in mind that in selecting a measure, one should always give careful consideration to the psychometric qualities of the instrument and the validation sample for the specific scale. For instance, scales developed with an adult eating disordered clinical sample may not necessarily be appropriate for use with a community sample of adolescents. In Table 4.1, we provide reliability data for the measures that are reviewed. The standard requirement for a scale to have acceptable reliability is a reliability coefficient of

(Text continues on page 126)

Table 4.1 Categorical Listing of Body Image Measures With Psychometric and Descriptive Information

Subjective and Affective Measures			
Measure	*Author(s)*	*Reliability*	*Description*
Body Dysmorphic Disorder Diagnostic Module	Phillips (1996)	Kappa: .96	Interview for BDD based on *DSM* criteria.
Body Dysmorphic Disorder Examination	Rosen and Reiter (1994)	IC: .81–.93; R: .87–.94 TR: .87-.94	Interview that assesses multiple components of BDD.
Body Esteem Scale-Children	Mendelson and White (1985); Cecil and Stanley (1997)	IC: (split half) = .85	Measures children's agreement with statements about one's body.
Body Esteem Scale–Adolescents and Adults	Mendelson, Mendelson, and White (2001)	IC: .81–.94	Twenty-three items assess three factors in adults and adolescents: appearance, attribution, and weight.
Body Image Assessment	Williamson, Gleaves, Watkins, and Schlundt (1993)	TR: current: .90, ideal: .71	Select current and ideal body shape from nine female figures that range from underweight to overweight.
Body Image Assessment–Children	Veron-Guidry and Williamson (1996)	TR (immediate): current = .94, ideal = .93; (1 week): current = .79, current/ideal = .67	Select current and ideal body shape from nine body figures that range from underweight to overweight.
Body Image Assessment–Revised	Beebe, Holmbeck, and Grzeskiewicz (1999)	TR: 2 weeks, cognitive = .74, affective = .79, desired = .70, cognitive-desired discrepancy = .67, affective-desired discrepancy = .79, affective-desired discrepancy = .63	Select cognitive and affective body estimates and desired body size from nine female figures that range from underweight to overweight. Normed with college females.

Measure	*Author(s)*	*Reliability*	*Description*
Body Image Coping Strategies Inventory	Cash, Santos, and Williams (2004); Cash and Grasso (2006)	Multiple studies: IC > .70	Twenty-nine items assess how individuals cope with various situations that threaten their body image on a 3-point Likert-type scale; three subscales (Avoidance, Appearance Fixing, Positive-Rational Acceptance).
Body Image Disturbance Questionnaire	Cash, Phillips, Santos, and Hrabosky (2004)	IC: .89	Items assess broad array of aspects of body image disturbance.
Body Image and Eating Questionnaire	Thelen, Powell, Lawrence, and Kuhnert (1992)	IC: all values ≥ .68	Fourteen items focus on overweight concerns, dieting, and restraint; for use with younger children.
Body Image Quality of Life Inventory	Cash and Fleming (2002)	IC = .95 TR: 2–3 weeks = .79	Assesses the impact of one's body image on 19 life domains using a 7-point Likert-type scale ranging from *very negatively* (–3) to *very positively* (+3).
Body Image Questionnaire	Huddy, Nieman, and Johnson (1993)	TR: 6 weeks: .97	Twenty items on a 3-point Likert scale from *agree* (1) to *disagree* (3).
Body Image States Scale	Cash, Fleming, and Alindogan (2002)	IC: females = .77 (neutral), .90 (negative), .88 (negative), .81 (positive), .80 (positive) IC: males = .62 (neutral), .66 (negative), .78 (negative), .83 (positive), .84 (positive) TR: females = .77, males = .72	Measures body image across five different contexts: neutral, negative (day at the beach and magazine models), and positive (party compliments and ideal weight).

(Continued)

Table 4.1 (Continued)

Measure	*Author(s)*	*Reliability*	*Description*
Body Mass Index Silhouette Matching Test	Peterson, Ellenberg, and Crossan (2003)	TR: males' current = .79, females' current = .85; males' ideal = .83, females' ideal = .82	Interval scale of 27 items along with four gender-specific figures to anchor for older children.
Body Parts Satisfaction Scale–Revised	Petrie, Tripp, and Harvey (2002)	IC: Study 1, body = .89, face = .76 IC: Study 2, body = .89, face = .74	Fourteen items assess satisfaction with various body sites on a 6-point Likert-type scale in adult women and men; two factors—body and face.
Body Satisfaction Scale	Siegel, Yancey, Aneshensel, and Schuler (1999)	IC: .73–.80	Rate satisfaction with four aspects of pubertal development; for males and females.
Body Shape Questionnaire	Cooper, Taylor, Cooper, and Fairburn (1987); Rosen, Jones, Ramirez, and Waxman (1996)	TR: .88	Thirty-four items assess adult men and women's concern with body shape and size.
Body Uneasiness Test (BUT)	Cuzzolaro, Vetrone, Marano, and Garfinkel (2006)	IC: BUT-A = .79–.90; BUT-B = .69–.90 TR: 1 week = BUT-A = .71–.91 (nonclinical); .80–.94 (clinical): BUT-B = .78–.94 (nonclinical); .68–.92 (clinical)	Thirty-four items assess body shape and/or weight dissatisfaction, behaviors, and feelings, and 37 items assess concern about specific body parts or functions; for males and females.
Breast/Chest Rating Scale	J. K. Thompson and Tantleff (1992)	TR: current = .85; ideal breast = .81, ideal chest = .69	Select current and ideal breast/chest ratings from five male and five female schematic figures, ranging from small to large upper torso.

Measure	*Author(s)*	*Reliability*	*Description*
Color-A-Person Dissatisfaction Test	Wooley and Roll (1991)	IC: .74–.85 TR: 2 weeks = .72–.84, 4 weeks = .75–.89	Uses five colors to indicate level of satisfaction with body sites by masking on a schematic figure for both men and women.
Contour Drawing Rating Scale	M. A. Thompson and Gray (1995); Wertheim, Paxton, and Tilgner (2004)	TR: 1 week, self = .79 TR: .65–.87	Select ideal and current self from nine male and nine female schematic figures, ranging from underweight to overweight. (Wertheim sample: adolescents).
Drive for Bulk	Furnham and Calnan (1998)	IC: .70	Rate desire to be bigger and to gain weight; for males between ages 16 and 18 years.
Drive for Muscularity Scale (DMS)	McCreary and Sasse (2000); McCreary, Sasse, Saucier, and Dorsch (2004)	IC: Full DMS-Males = .85, MBI-Males = .88; MB-Men = .81; Full DMS-Females = .82	Fourteen items assess drive for muscularity: Muscle-Oriented Body Image (MBI; seven items), Muscle-Oriented Behavior (MB; seven items).
Drive for Muscularity–Body Image subscale	(1) McCreary and Sasse (2000) (2) Cafri, van den Berg, and Thompson (2006)	(1) IC = .84 (whole scale) (2) IC = .90 (subscale)	Seven items assess satisfaction with appearance (specifically muscular physique); adolescents and young adults.
Eating Disorder Inventory-2 Body Dissatisfaction subscale	Garner (1991)	IC >.70 for multiple samples, including individuals with eating disorders and non-eating-disordered controls	Nine items assess dissatisfaction with specific body parts.

(Continued)

Table 4.1 (Continued)

Measure	*Author(s)*	*Reliability*	*Description*
Feelings of Fatness Questionnaire	Roth and Armstrong (1993)	IC: Troubles = .96, Satisfaction = .98	Measures the extent to which females feel thin or fat across 61 situations; two subscales: Troubles (38 items) and Satisfaction (23 items).
Figure Rating Scale	Stunkard, Sorenson, and Schulsinger (1983); J. K. Thompson and Altabe (1991)	TR: 2 weeks: ideal; males = .82, females = .71, Self-think: males = .92, females = .89, Self-feel; males = .81; females = .83	Select ideal and current self from nine male and female figures that vary from underweight to overweight. Normed with college males and females.
Goldfarb Fear of Fat Scale	Goldfarb, Dykens, and Gerrard (1985)	IC: .70; TR = .88	Ten items assess overconcern with fatness and body size.
Kid's Eating Disorder Survey	Childress, Brewerton, Hodges, and Jarrell (1993)	TR (4 months): .83 for entire survey; not given for figures only	Choose ideal and current self from eight male and eight female figure drawings: for preadolescent childres, Grades 5 to 8. Items relate to weight control behaviors.
Male Body Attitudes Test	Tylka, Bergeron, and Schwartz (2005)	IC: .80–.94; TR: .81–.94	Twenty-four items assess body image related to muscularity, body fat, body shape, and height.

Measure	*Author(s)*	*Reliability*	*Description*
McKnight Risk Factor Survey (MFRS-III)	Shisslak et al. (1999)	IC: elementary = .82; middle school = .86; high school = .87 TR: elementary = .79; middle school = .84; high school = .90	Five items assess concern with weight and shape; elementary, middle and high school children.
Multidimensional Body-Self Relations Questionnaire–Appearance Evaluation subscale	Cash (body-images.com); Brown, Cash, and Mikulka (1990)	IC > .70 for multiple samples, including men and women	Seven items assess overall appearance satisfaction.
Muscle Appearance Satisfaction Scale	Mayville, Williamson, White, Netemeyer, and Drab (2002)	IC: .74–.79 (subscales); TR: .76–.89 (subscales)	Nineteen items assess excessive concern with muscularity in males; five subscales (bodybuilding dependence, muscle checking, substance use, injury, and muscle satisfaction).
Muscle Dysmorphic Disorder Inventory	Hildebrandt, Langenbucher, and Schlundt (2004)	IC: .77–.85; TR: .81–.87	Thirteen items; three subscales (Desire for Size, Appearance Intolerance, Functional Impairment) in males using a 5-point Likert-type scale.
Physical Appearance State and Trait Anxiety Scale (PASTAS)	Reed, Thompson, Brannick, and Sacco (1991)	IC: trait = .82–.88, state = .82–.92 TR: 2 weeks = .87	Assesses trait or state anxiety associated with 16 body sites (8 weight related, 8 non–weight related).
Self-Image Questionnaire for Young Adolescents–Body Image subscale	Petersen, Schulenberg, Abramowitz, Offer, and Jarcho (1984)	IC: boys = .81, girls =.77 TR: 1 year = .60; 2 years = .44	Eleven items assess positive feelings toward the body in 10- to 15-year-old boys and girls.

(Continued)

Table 4.1 (Continued)

Measure	*Author(s)*	*Reliability*	*Description*
Situational Inventory of Body Image Dysphoria (SIBID)–Short Form	Cash (2002)	IC: females = .94–.96; males = .93–.94 TR: 1 month, females = .87; males = .81	Twenty items; short form of SIBID. Men and women rate how often they experience negative body image emotions in various situations.
Somatomorphic Matrix	(1) Gruber, Pope, Borowiecki, and Cohane (1999) (2) Cafri, Roehrig, and Thompson (2004)	(1) Not available (2) TR: females = .35–.75; males = .45–.80	Assesses current and ideal muscularity and body fat dimensions in men and women; available in (1) computer and (2) paper-and-pencil forms.
Cognitive Measures			
Measure	*Author(s)*	*Reliability*	*Description*
Appearance Schemas Inventory–Revised	Cash, Melnyk, and Hrabosky (2004)	IC = total-women = .88; total-men = .90; self-eval-women = .82; self-eval-men = .84; motivational-women = .90, motivational-men = .91	Twenty items assess beliefs and assumptions about the importance and influence of appearance in one's life on a 5-point Likert-type scale; two subscales (Self-Evaluative Saliance [12 items], Motivational Saliance [8 items]); for males and females.
Assessment of Body Image Cognitions	Jakatdar, Cash, and Engle (2006)	IC > .70	Thirty-seven items assess eight types of distorted thinking about own physical appearance.
Beliefs About Appearance Scale	Spangler and Stice (2001)	IC: .94, .95, .96 (three samples, respectively) TR: 3 weeks = .83 (Sample 3); 10 months = .73 (Sample 2)	Twenty items assess one's beliefs about consequences of appearance for relationships, achievement, self-view, and feelings on a 5-point Likert-type scale.

Measure	*Author(s)*	*Reliability*	*Description*
Body Attitude Test (BAT)–Japanese version	Kashima, Yamashita, and Okamoto (2003)	IC: .90	Twenty items measure body dissatisfaction.
Body Image Automatic Thoughts Questionnaire	Cash, Lewis, and Keeton (1987)	IC: .90	Measures frequency of 37 negative and 15 positive body image cognitions in women.
Body Image Ideals Questionnaire	Cash and Szymanski (1995)	IC: discrepancy = .75; importance = .82	Ratings of ideal and actual self on 10 attributes related to weight/appearance and strength/importance of attribute.
Body Image Ideals Questionnaire–Expanded	Szymanski and Cash (1995)	IC: .81–.95	Ratings of one's specific attributes from own viewpoint and that of romantic partner based on "ideal" and "actual."
Bulimia Cognitive Distortions Scale–Physical Appearance subscale	Schulman, Kinder, Powers, Prange, and Gleghorn (1986)	IC: .97 (entire scale)	Measures agreement with 25 statements related to physical appearance-related cognitions.
Drive for Muscularity Attitudes Questionnaire	Morrison, Morrison, Hopkins, and Rowan (2004)	IC: .84 (Study 1), .82 (Study 2); .79 (protein supplement), .72 (weight training), .45 (cardio exercise)	Eight items assess attitudes toward muscularity in males.
Eating Disorder Belief Questionnaire–Self Acceptance & Acceptance by Others subscales	Rose, Cooper, and Turner (2006)	IC: .85 (self-acceptance), .94 (acceptance by others)	Six items assess how attractive one will feel if one is slim (self-acceptance); nine items assess how others will feel if body is slim/toned (acceptance by others).
Swansea Muscularity Attitudes Questionnaire	Edwards and Launder (2000);	IC: .86–.92; .58–.70 (for 2-item subscale)	Nineteen items assess intention to become more muscular (eight items), positive attributes of muscularity (nine items), and muscle-building activities (two items).

(Continued)

Table 4.1 (Continued)

Behavioral Measures			
Measure	*Author(s)*	*Reliability*	*Description*
Body Checking Questionnaire	Reas, Whisenhunt, and Netemetyer (2002)	IC: .88 (overall appearance), .92 (specific body parts), .83 (idiosyncratic checking) TR: 2 weeks = .94	Twenty-three items assess body-checking behaviors related to overall appearance, specific body parts, and idiosyncratic checking behaviors in adult females.
Body Image Avoidance Questionnaire	Rosen, Srebnik, Saltzberg, and Wendt (1991)	IC: .89; TR: 2 weeks = .87	Frequency of body image avoidance behaviors.
Sociocultural and Interpersonal Influences			
Measure	*Author(s)*	*Reliability*	*Description*
Body Comparison Scale	Fisher and Thompson (1998)	IC: .95	Assesses tendency to compare specific body sites with others.
Fear of Negative Appearance Evaluation Scale	Lundgren, Anderson, and Thompson (2004)	IC = .94	Measures affective reaction to perceived negative evaluation by others. The original measure consisted of eight items. In the second study, it was reduced it to 6 items.
Feedback on Physical Appearance Scale	Tantleff-Dunn, Thompson, and Dunn (1995)	IC: .84 TR: 2 weeks = .82	Twenty-six items assess verbal and nonverbal appearance-related commentary on a 5-point Likert-type scale; adults.
Ideal Body Internalization Scale–Revised	Stice, Shaw, and Nemeroff (1998)	IC: .88 TR: 1 year = .59	Ten items assess agreement with cultural ideal for female body.

Measure	*Author(s)*	*Reliability*	*Description*
Objectified Body Consciousness Scale	McKinley and Hyde (1996); Lindberg, Hyde and McKinley (2006)	IC: body surveillance: .79–.89; body shame = .75	Sixteen items assess body surveillance (eight items) and body shame (eight items); modified version used with preadolescents and adolescents.
Perception of Teasing Scale	J. K. Thompson, Cattarin, Fowler, and Fisher (1995)	IC: general weight = .94, competency = .78	Twelve items assess frequency and emotional response to general weight teasing and competency teasing.
Physical Appearance Comparison Scale	J. K. Thompson, Heinberg, and Tantleff (1991)	IC = .78; TR = .72	Assesses tendency to compare own appearance with others.
Self-Objectification Questionnaire	Noll and Fredrickson (1998)	NA	Rank order of 10 attributes (5 appearance-based, 5 non–appearance based) based on how important the attribute is to the physical self-concept.
Sociocultural Attitudes Towards Appearance Scale–3	J. K. Thompson, van den Berg, Roehrig, Guarda, and Heinberg. (2004)	IC: .93	Thirty items measure multiple aspects of societal influence on appearance and consists of four distinct subscales: Pressures, Information, Internalization-General, Internalization-Athlete.
Sociocultural Internalization of Appearance Questionnaire–Adolescents	Keery, Shroff, Thompson, Wertheim, and Smolak (2004)	IC: .83–.92 (U.S. and international samples)	Five items measure thin-ideal internalization in adolescent girls.

NOTE: BDD = body dysmorphic disorder; IC = internal consistency; TR = test-retest reliability; NA = not available.

around .70 (J. K. Thompson, 2004). When a measure is considered for clinical or research purposes, it is important to select a measure that has acceptable reliability and validity and has been evaluated on a sample that is similar to the one under consideration. Some measures have received limited validity work, but others have received extensive validation (such as the Eating Disorders Inventory–Body Dissatisfaction scale and the Multidimensional Body Self-Relations Questionnaire; see Cash, 2000; Garner, 2004). In the sections that follow, we discuss measures that fit into widely accepted categorical dimensions of body image, with an intent of providing a broad overview of the possible measures that might be included in an assessment battery. Again, this is a highly selective review, and the reader is also encouraged to examine the broader range of potential measures that are available in Table 4.1.

Measurement Categories

Subjective and Affective Measures

Subjective and affective measures tap into a dimension of body image that is usually labeled *satisfaction* (J. K. Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999). The satisfaction may be site specific (e.g., waist, hips) or may involve a more global self-evaluation of overall appearance. There are two primary methods for assessing the subjective and affective component. One method consists of using figural stimuli consisting of a range of images (usually schematic line drawings or silhouettes) that differ in body size and/or shape. The stimuli are presented to the participants, who are asked to select the image that represents their ideal self and their real (current) self. The discrepancy between their current and ideal selves is used as the index of dissatisfaction. The second category consists of questionnaire measures that typically ask respondents to complete a variety of items designed to assess body satisfaction or an affective dimension of body image. Examples of these two types of measurement strategies (figural rating scales, questionnaire measures) are provided below.

A variety of figural stimuli measures have been created in the past two decades (Table 4.1). A measure that has been developed specifically for children is the Body Image Assessment–Children (Veron-Guidry & Williamson, 1996). This scale is particularly useful as it has separate figures for children and for adolescents, so that the participants make their ratings using figures similar to themselves in body composition and pubertal status. A figural scale that has been used with adult women is the Contour Drawing Rating Scale (J. K. Thompson et al., 1999; M. A. Thompson & Gray, 1995). Two widely used figural scales allow for the assessment of dissatisfaction related to overall weight and muscularity. The Somatomorphic Matrix is a two-dimensional computerized body image test that can be used to assess self and ideal body image in relation to muscularity and body fat (Gruber, Pope, Borowiecki, & Cohane, 1999). The test is available for both genders and consists of 100 images arranged in a 10×10 matrix, representing 10 degrees of adiposity and 10 degrees of muscularity. The Somatomorphic Matrix Modification is a paper-and-pencil modification of the original somatomorphic matrix (Cafri & Thompson, 2004c). The Bodybuilder Image Grid (Hildebrandt, Langenbucher, & Schlundt, 2004) is a paper-and-pencil measure that has figures arranged in a 6×5 matrix with body fat and muscularity varying on two separate dimensions (i.e., six degrees of body fat and five degrees of muscularity).

With regard to questionnaire measures, there is also a wide variety available to assess the subjective and/or affective component of body image. Some of these measures are global or generic in nature and provide an overall estimate of body dissatisfaction or focus on specific sites. The Self Image Questionnaire for Young Adults (Petersen, Schulenberg, Abramowitz, Offer, & Jarcho, 1984) includes a subscale that provides a global measure of body satisfaction for adolescents. The Appearance Evaluation subscale of the Multidimensional Body Self-Relations Questionnaire (Cash, 2000) is a measure that has been used in numerous studies, with adult

men and women of different ethnicities (see www.body-images.com). A sample item from this scale is "I like my looks just the way they are." The Body Shape Questionnaire (Cooper, Taylor, Cooper, & Fairburn, 1987) assesses the experience of "feeling fat" with items such as, "Have you worried about your flesh not being firm enough?"

Other measures offer site-specific information regarding satisfaction or body image affect, such as the Body Uneasiness Test (Cuzzolaro, Vetrone, Marano, & Garfinkel, 2006), created for adolescents and adults, which has items that assess concern about specific body parts. Likewise, the Body Dissatisfaction subscale of the Eating Disorder Inventory–3 (Garner, 2004) asks participants to rate satisfaction with nine weight-relevant body sites, including the stomach, hips, thighs, and so forth. This scale has also been widely used with adults and adolescents and has a wealth of normative information for use with individuals with eating disorders. The Body Parts Satisfaction Scale (Petrie, Tripp, & Harvey, 2002) also assesses satisfaction with various body sites in adult men and women.

Some scales assess a more specific affective component of body image. For instance, the Physical Appearance State and Trait Anxiety Scale (Reed, Thompson, Brannick, & Sacco, 1991) provides for an assessment of anxiety regarding weight and non-weight-related (hair, eyes, face) body aspects. Cash's (1994, 2002) Situational Inventory of Body Image Dysphoria scale taps into negative body image–related emotions that accompany exposure to various situations.

Cognitive Measures

Scales assessing cognitive components of body image include those that evaluate thoughts or beliefs about one's appearance. For example, Cash, Melnyk, and Hrabosky's (2004) Appearance Schema Inventory–Revised assesses beliefs about appearance that may reflect a rather ingrained aspect of body image schema. The Body Image Automatic Thoughts Questionnaire (Cash, Lewis, & Keeton, 1987) has subscales that assess Positive and Negative body-related cognitions. A relatively new and somewhat unique measure is the Eating Disorder Belief Questionnaire, which contains the Self-Acceptance subscale and the Acceptance by Others subscale (Rose, Cooper, & Turner, 2006). The subscales assess how an individual will feel if her or his body is toned and how others will feel if her or his body is toned. Within this category are a variety of measures related to appearance ideals and self-schema (e.g., Stein & Hedger, 1997; Szymanski & Cash, 1995).

Behavioral Assessment

Behavioral measures attempt to document the specific observational manifestation of body image disturbance, such as avoidance or body checking. One of the first measures of this type was a self-report measure developed by Rosen, Srebnik, Saltzberg, and Wendt (1991), the Body Image Avoidance Questionnaire. The Body Checking Questionnaire (Reas, Whisenhunt, & Netemeyer, 2002) is one of the more recent innovations in this area; a 23-item scale that indexes a variety of body-checking behaviors. Although only a few questionnaire measures specifically focus on behavioral issues, some of the measures of body dysmorphic disorder (see Table 4.1) provide an assessment of the behavioral component. Interestingly, most of the measures in this category provide a self-report of one's behaviors—there remains a need for further development of objective measures of behavioral avoidance.

Interview Scales

In addition to the above-mentioned methods of assessment, interview strategies have been used with children, adolescents, and adults. The Eating Disorder Examination (EDE; Fairburn & Cooper, 1993) is a semi-structured interview of eating pathology currently in its 12th edition. The interview has subscales related to body

image (shape and weight concerns). Two interviewers provide symptom ratings in this assessment tool. The interview was originally designed for adults, but Bryant-Waugh, Cooper, Taylor, and Lask (1996) modified the EDE and administered it to children and found that the measure works well with children. The Structured Interview for Anorexia and Bulimic Disorders (SIAB) also has a body image scale as one of its six factors (Fichter, Herpertz, Quadflieg, & Herpertz-Dahlmann, 1998). The Interview for Diagnosis of Eating Disorders–IV (Kutlesic, Williamson, Gleaves, Barbin, & Murphy-Eberenz, 1998) includes symptom ratings for body image disturbance, as well as the other *DSM* criteria for eating disorders (anorexia nervosa, bulimia nervosa, binge eating disorder, and eating disorder not otherwise specified).

Body Dysmorphic Disorder

Body dysmorphic disorder (BDD) is an extreme form of body image disturbance wherein the disparagement of a particular aspect of appearance may be very severe, even delusional. In recent years, a variety of measures have been developed to assess this disorder. For adolescents, the Body Image Rating Scale (BIRS) has been designed to assess the presence and severity of BDD and its associated features (Mayville, Katz, Gipson, & Cabral, 1999). The scale consists of 15 items in a Likert-type format that focuses on cognitive, affective, and behavioral features of BDD. The most widely used interview measure of BDD is the Body Dysmorphic Disorder Examination (BDDE; Rosen & Reiter, 1996), which has 34 items that index symptoms of BDD. The content of the interview includes preoccupation/negative evaluation of appearance, self-consciousness, excessive importance given to appearance in self-evaluation, avoidance of social situations or activities, camouflaging appearance, and body-checking behavior.

The Body Dysmorphic Disorder Diagnostic Module (BDDDM) is an interview based on *DSM-IV* diagnostic criteria that is designed similar to other Structured Clinical Interview for DSM Disorders (SCID) modules to determine whether a diagnosis of body dysmorphic disorder is appropriate (Phillips, 1996). The BDD modification of the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) is a 12-item interview that assesses severity of BDD symptoms (Phillips, 1996). These items assess BDD-related thoughts, behaviors, insight, and avoidance.

Within the field of BDD, Pope and colleagues have identified a clinical disorder that corresponds to a pathological preoccupation with the pursuit of a muscular ideal, referred to as *muscle dysmorphia* (Pope, Katz, & Hudson, 1993; Pope, Gruber, Choi, Olivardia, & Phillips, 1997). In muscle dysmorphia, a person is described as experiencing cognitive symptoms, including extreme body dissatisfaction and repeated thoughts of not being sufficiently muscular, along with behavioral symptoms such as substance abuse (e.g., use of anabolic steroids), strict dieting, compulsive weight lifting, and mirror checking (Olivardia, 2004). Some of the measures included in Table 4.1 that assess a muscularity component of body image (Drive for Muscularity, Muscle Areas Satisfaction Scale, etc.) are also useful in providing information related to diagnosing muscle dysmorphia.

Recent Advances

One of the recent trends in body image assessment work is the development of a variety of measures designed to index constructs that are closely related to body image and/or that might be considered risk factors for the development of body image disturbance. For instance, the Body Image Coping Strategies Inventory (Cash, Santos, & Williams, 2004) looks at how individuals cope with situations that threaten their body image. The Body Image Quality of Life Inventory (Cash & Fleming, 2002) asks respondents about the effect of their body image on a variety of life domains (home, work, etc.).

A variety of sociocultural and interpersonal factors, such as media influences and psychosocial pressures (e.g., teasing), have been associated with body image disturbance. Utilization of these

measures may be useful not only for understanding factors that are connected to an individual's body image but also assist in identifying variables that may be linked to disturbed eating patterns (Levine & Harrison, 2004; J. K. Thompson, van den Berg, Roehrig, Guarda, & Heinberg, 2004). One of the most commonly used measures in this category is the Sociocultural Attitudes Towards Appearance Questionnaire–Third Revision (J. K. Thompson et al., 2004), which measures multiple aspects of societal influence on appearance, including four subscales: Pressures, Information, Internalization-General, Internalization-Athlete. This scale has also been validated for use with clinical eating disorders populations (Calogero, Davis, & Thompson, 2004).

Internalization measures that examine the extent to which the "thin ideal" present in the media is internalized by an individual are also in this category of measures (e.g., Sociocultural Internalization of Appearance Questionnaire–Adolescents [Keery, Shroff, Thompson, Wertheim, & Smolak, 2004]; Ideal Body Internalization Questionnaire [Stice, Shaw, & Nemeroff, 1998]). Appearance comparison scales (e.g., the Physical Appearance Comparison Scale [J. K. Thompson, Heinberg, & Tantleff, 1991]) assess the degree to which an individual compares himself or herself with peers, models, and others. Research with these types of measures are frequently used in studies evaluating risk for body image and eating disturbances (e.g., Shroff & Thompson, 2006a, 2006b).

Recently developed measures of body objectification are also relevant to the measurement of body image. One commonly used measure in this category is the Objectified Body Consciousness Scale (OBC; McKinley & Hyde, 1996), which includes two subscales related to how individuals respond to cultural pressures to meet appearance ideals. The Body Surveillance subscale assesses the degree to which an individual engages in habitual monitoring of their external appearance. The Body Shame subscale assesses the degree to which an individual feels like a bad person if not meeting societal appearance ideals. The OBC scale has been modified recently to provide a validated measure of objectified body consciousness in preadolescents and adolescents (Lindberg, Hyde, & McKinley, 2006). The Self-Objectification Questionnaire (Noll & Fredrickson, 1998) provides a short rank-order scale that is used to measure the extent to which an individual values observable appearance-based attributes over nonobservable competence-based attributes as most important to their physical self-concept. Research with these measures has frequently used them to assess the extent to which individuals view their bodies as objects to be looked at and evaluated by others, which, in turn, has been linked to an increased risk for body image and eating disturbances (Calogero, Davis, & Thompson, 2005; Tiggemann & Lynch, 2001; Tiggemann & Slater, 2001).

Researchers are also beginning to realize the value of cross-cultural studies of body image. To that end, several studies have examined the relevance and applicability of translating measures created in English into other languages. For example, Kashima, Yamashita, and Okamoto (2003) have translated the Body Attitude Test into Japanese, and Thurfjell, Edlund, Arinell, Hagglof, and Engstrom (2003) have created a Swedish translation of the Body Dissatisfaction subscale of the Eating Disorders Inventory for Children. Both of these scales show good internal consistency in the samples in which they were studied. Rousseau, Knotter, Barbe, Raich, and Chabrol (2005) have used a French version of the Body Shape Questionnaire with university female students. In addition, Lau, Lum, Chronister, and Forrest (2006) have translated the Body Parts Satisfaction Scale and the Body Comparison Scale (Fisher & Thompson, 1998) into Chinese. Yamamiya, Shroff, and Thompson (2008) translated a variety of measures into Japanese. These new measures will allow researchers to assess the same construct across cultures and to perform comparative studies. Such studies may shed some light on factors leading to body image disturbance in various cultures around the world.

Methodological Issues

An important consideration about body image scales is whether the clinician or researcher is

attempting to assess a state or trait dimension. In most cases, measures assess traits; however, a few state measures have been developed. Primarily, these measures have been used in experimental studies designed to investigate the impact of specific variables (e.g., exposure to media images) on body image (J. K. Thompson, 2004). These scales could potentially be used in a clinical setting as a gauge of the immediate, within-session improvement in body image. One such measure is the state subscale of the Physical Appearance State and Trait Anxiety Scale (Reed et al., 1991). Cash, Fleming, and Alindogan (2002) have also created the Body Image States scale, which offers the advantage of being rather short (six items) so that it can be used efficiently for multiple ratings.

Another methodological issue, touched on early in this chapter, is that some of the measures created for the assessment of some component of body image were standardized on relatively small samples and may have limited evidence of reliability and validity. In addition, the samples are not always clinical in nature (i.e., individuals with a diagnosis of an eating disorder or body dysmorphic disorder). As mentioned earlier, they may have limited generalizability due to limited work with a variety of ages and/or ethnicities. A further limitation of body image measures is that no reliable measures for very young children (i.e., younger than 6 years) have been developed.

The ethnicity of the standardization sample should be taken into account. For example, using a figural rating scale, a sample of Moroccan Sahraoui women rated their ideal body size as significantly *larger* than their rating of an average body size (Rguibi & Belahsen, 2006), which is in contrast to the typical findings among White, European American samples of women. In this sample, greater body dissatisfaction predicted attempts to gain weight as opposed to lose weight. Ethnic differences in response to body image measures have also been observed based on the degree of assimilation and/or acculturation into other cultural contexts. For example, again using figural rating scales, research has indicated that Latina women born in the United States endorse an even thinner ideal body size than European American women, whereas Latina women who immigrated to the United States endorsed a larger body ideal (Lopez, Blix, & Blix, 1995). Although limited work has assessed levels of different body image dimensions with a variety of ethnic groups, it is important to consider ethnic differences when interpreting individual responses to body image measures and to seek out this information when it is available.

The gender and sexual orientation of the standardization sample should also be taken into account. For example, the body objectification measures have been primarily developed for and tested in samples of heterosexual women; however, recent research has provided more information about men's body surveillance and body shame (Hebl, King, & Lin, 2004), as well as gay men (Martins, Tiggemann, & Kirkbride, 2007) and lesbian women (Kozee & Tylka, 2006), indicating different levels and patterns of relations among body objectification and body image disturbances across these populations.

Conclusions and Practical Considerations

We now offer some specific guidelines for selecting a measure. One of the strategies that has guided our decision over the past several years (e.g., J. K. Thompson et al., 1999) is to include a variety of measures that tap into different dimensions of body image disturbance, especially when involved in clinical work. (The decision for research purposes may be more specific and targeted, depending on the research question.) For instance, we generally include at least one measure that assesses a more abstract aspect of appearance (e.g., not weight related or specific to a certain body site). One widely used measure of this aspect of body image is the Multidimensional Body Self-Relations Questionnaire–Appearance Evaluation subscale (MBSRQ-AE; Brown, Cash, & Mikulka, 1990). We also typically include a measure of weight-specific body image, such as the Eating Disorder

Inventory–Body Dissatisfaction subscale. Another measure that might be included is a measure that assesses a variety of different sites, such as the MBSRQ–Body Areas Satisfaction Scale. A nice addition to questionnaire measures is a figural rating scale, such as the Contour Drawing Rating scale (M. A. Thompson & Gray, 1995).

The above measures generally index the affective and subjective nature of body image; therefore, it may also be useful to add a measure or two that involves cognitive disturbances (see Table 4.1). These types of measures are very helpful for providing evidence of disturbed thinking patterns that might be addressed in a clinical intervention for body image or related disturbances (e.g., eating disorders). For instance, the Appearance Schemas Inventory–Revised (Cash, Melnyk, & Hrabosky, 2004) and the Assessment of Body Image Cognitions (Jakatdar, Cash, & Engle, 2006) are two measures that appear to have excellent psychometric properties.

Especially when dealing with boys or men or an athletic sample, it might be important to include a measure that allows for the specific assessment of muscularity dissatisfaction. As Table 4.1 shows, many of these measures have been developed in recent years, including the Drive for Muscularity Scale (McCreary, Sasse, Saucier, & Dorsch, 2004), Drive for Muscularity Attitudes Questionnaire (Morrison, Morrison, Hopkins, & Rowan, 2004), and others (for a review, see Cafri & Thompson, 2007).

If an initial assessment suggests the possible presence of body dysmorphic disorder, it would be important to include a measure that assesses this more severe form of disturbance such as the Body Dysmorphic Disorder Examination (Rosen & Reiter, 1996) or the Body Dysmorphic Disorder Diagnostic Module (Phillips, 1996).

There are a few sources that someone interested in the assessment of body image can return to periodically for updates on measures and assessment issues. First, the journal *Body Image* (Elsevier.com) frequently has articles that include development and validation of new measures. Second, Tom Cash's Web site (body-images.com) contains detailed advice on his measures of body image, which are quite numerous. Third, our Web site (bodyimagedisturbance.org) contains information on the measures that we have developed that assess body image and/or related constructs (such as social and interpersonal influences on body image). Fourth, it is very important to stay current with the scientific literature because new measures appear with regularity and old measures are reexamined. For instance, Pook, Tuschen-Caffier, and Brähler (2008) recently evaluated eight different versions of the Body Shape Questionnaire. Measures are also often modified and evaluated for use with samples cross-culturally; for instance, Yamamiya et al. (2008) recently evaluated several measures on a Japanese sample. Therefore, when considering a measure that has been available for several years, it is often useful to conduct a literature review to see if the measure has been modified and/or whether the measure has received new evaluation with a sample that it may not have been tested upon previously.

In conclusion, this chapter has provided a fairly selective review of measures commonly used in the assessment of body image. The clinician or researcher has a rather daunting task when faced with the variety and plethora of measures available for selection. We hope this chapter offers a few guidelines that reduce the complexity involved in the evaluation of body image disturbance.

References

Beebe, D. W., Holmbeck, G. N., & Grzeskiewicz, C. (1999). Normative and psychometric data on the Body Image Assessment–Revised. *Journal of Personality Assessment, 73,* 374–394.

Brown, T. A., Cash, T. F., & Mikulka, P. J. (1990). Attitudinal body-image assessment: Factor analysis of the Body Self Relations Questionnaire. *Journal of Personality Assessment, 55,* 135–144.

Bryant-Waugh, R. J., Cooper, P. J., Taylor, C. L., & Lask, B. D. (1996). The use of the Eating Disorder Examination with children: A pilot study. *International Journal of Eating Disorders, 19,* 391–397.

Cafri, G., Roehrig, M., & Thompson, J. K. (2004). Reliability assessment of the somatomorphic matrix. *International Journal of Eating Disorders, 35,* 597–600.

Cafri, G., & Thompson, J. K. (2004a). *Development of a modified version of the somatomorphic matrix.* Unpublished measure.

Cafri, G., & Thompson, J. K. (2004b). Evaluating the convergence of muscle appearance attitude measures. *Assessment, 11,* 224–229.

Cafri, G., & Thompson, J. K. (2004c). Measuring male body image: A review of the current methodology. *Psychology of Men & Masculinity, 5,* 18–29.

Cafri, G., & Thompson, J. K. (2007). Measurement of the muscular ideal. In J. K. Thompson & G. Cafri (Eds.), *The muscular ideal.* Washington, DC: American Psychological Association.

Cafri, G., van den Berg, P., & Thompson, J. K. (2006). Pursuit of muscularity in adolescent boys: Relations among biopsychosocial variables and clinical outcomes. *Journal of Clinical Child and Adolescent Psychology, 35,* 283–291.

Calogero, R. M., Davis, W. N., & Thompson, J. K. (2004). The Sociocultural Attitudes Toward Appearance Scale (SATAQ-3): Reliability and normative data for eating disorder patients. *Body Image: An International Journal of Research, 1,* 193–198.

Calogero, R. M., Davis, W., N., & Thompson, J. K. (2005). The role of self-objectification in the experience of women with eating disorders. *Sex Roles, 52,* 43–50.

Cash, T. F. (1994). The Situational Inventory of Body-Image Dysphoria: Contextual assessment of a negative body image. *The Behavior Therapist, 17,* 133–134.

Cash, T. F. (2000). *The users' manuals for the Multidimensional Body-Self Relations Questionnaire, Situational Inventory.* Retrieved September 2007 from Body-images.com

Cash, T. F. (2002). The Situational Inventory of Body-Image Dysphoria: Psychometric evidence and development of a short form. *International Journal of Eating Disorders, 32,* 362–366.

Cash, T. F., & Fleming, E. C. (2002). The impact of body image experiences: Development of the Body Image Quality of Life Inventory. *International Journal of Eating Disorders, 31,* 455–460.

Cash, T. F., Fleming, E. C., & Alindogan, J. (2002). Beyond body image as a trait: The development and validation of the Body Image States Scale. *Eating Disorders: The Journal of Treatment and Prevention, 10,* 103–113.

Cash, T. F., & Grasso, K. (2005). The norms and stability of new measures of the multidimensional body image construct. *Body Image, 2,* 199–203.

Cash, T. F., & LeBarge, A. S. (1996). Development of the Appearance Schemas Inventory: A new cognitive body image assessment. *Cognitive Therapy and Research, 20,* 37–50.

Cash, T. F., Lewis, R. J., & Keeton, P. (1987, March). *Development and validation of the Body-Image Automatic Thoughts Questionnaire: A measure of body-related cognitions.* Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, GA.

Cash, T. F., Melnyk, S. E., & Hrabosky, J. I. (2004). The assessment of body image investment: An extensive revision of the Appearance Schemas Inventory. *International Journal of Eating Disorders, 35,* 305–316.

Cash, T. F., Phillips, K. A., Santos, M. T., & Hrabosky, J. I. (2004). Measuring "negative body image": Validation of the Body Image Disturbance Questionnaire in a nonclinical population. *Body Image, 1,* 363–372.

Cash, T. F., & Pruzinsky, T. (2002). Future challenges for body image theory, research, and clinical, practice. In T. F. Cash & T. Pruzinsky (Eds.), *Body images: A handbook of theory, research, and clinical practice* (pp. 509–516). New York: Guilford.

Cash, T. F., Santos, M., & Williams, E. (2004). *Coping with body-image threats and challenges: Validation of the Body Image Coping Strategies Inventory.* Unpublished manuscript.

Cash, T. F., & Szymanski, M. L. (1995). The development and validation of the Body-Image Ideals Questionnaire. *Journal of Personality Assessment, 64,* 466–477.

Cecil, H., & Stanley, M. A. (1997). Reliability and validity of adolescents' scores on the Body Esteem Scale. *Educational and Psychological Measurement, 57,* 340–356.

Childress, A. C., Brewerton, T. D., Hodges, E. L., & Jarrell, M. P. (1993). The Kids' Eating Disorders Survey (KEDS): A study of middle school students. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 843–850.

Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairburn, C. G. (1987). The development and validation of the Body Shape Questionnaire. *International Journal of Eating Disorders, 6,* 485–494.

Cuzzolaro, M., Vetrone, G., Marano, G., & Garfinkel, P. E. (2006). The Body Uneasiness Test (BUT): Development and validation of a new body image assessment scale. *Eating and Weight Disorders, 11,* 1–13.

Edwards, S., & Launder, C. (2000). Investigating muscularity concerns in male body image: Development of the Swansea Muscularity Attitudes Questionnaire. *International Journal of Eating Disorders, 28,* 120–124.

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

Fichter, M. M., Herpertz, S., Quadflieg, N., & Herpertz-Dahlmann, B. (1998). Structured interview for anorexic and bulimic disorders for DSM-IV and ICD-10: Updated (third) revision. *International Journal of Eating Disorders, 24,* 227–249.

Fisher, E., & Thompson, J. K. (1998). *Social comparison and body image: An investigation of body comparison processes using multidimensional scaling.* Unpublished manuscript.

Furnham, A., & Calnan, A. (1998). Eating disturbance, self-esteem, reasons for exercising and body weight dissatisfaction in adolescent males. *European Eating Disorders Review, 6,* 58–72.

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

Garner, D. (2004). *Manual for the Eating Disorders Inventory–3 (EDI-3).* Odessa, FL: Psychological Assessment Resources.

Goldfarb, L. A., Dykens, E. M., & Gerrard, M. (1985). The Goldfarb Fear of Fat Scale. *Journal of Personality Assessment, 49,* 329–332.

Gruber, A. J., Pope, H. G., Borowiecki, J., & Cohane, G. (1999). The development of the somatomorphic matrix: A bi-axial instrument for measuring body image in men and women. In T. S. Olds, J. Dollman, & K. I. Norton (Eds.), *Kinanthropometry VI* (pp. 217–231). Sydney, Australia: International Society for the Advancement of Kinanthropometry.

Hebl, M. R., King, E. B., & Lin, J. (2004). The swimsuit becomes us all: Ethnicity, gender, and vulnerability to self-objectification. *Personality and Social Psychology Bulletin, 30,* 1322–1331.

Hildebrandt, T., Langenbucher, J., & Schlundt, D. G. (2004). Muscularity concerns among men: Development of attitudinal and perceptual measures. *Body Image, 1,* 169–181.

Huddy, D. C., Nieman, D. C., & Johnson, R. L. (1993). Relationship between body image and percent body fat among college male varsity athletes and nonathletes. *Perceptual and Motor Skills, 77,* 851–857.

Jakatdar, T. A., Cash, T. F., & Engle, E. K. (2006). Body-image thought processes: The development and initial validation of the Assessment of Body-Image Cognitive Distortions. *Body Image, 3,* 325–333.

Kashima, A., Yamashita, T., & Okamoto, A. (2003). Japanese version of the Body Attitude Test: Its reliability and validity. *Psychiatry and Clinical Neurosciences, 57,* 511–516.

Keery, H., Shroff, H., Thompson, J. K., Wertheim, E., & Smolak, L. (2004). The Sociocultural Internalization of Appearance Scale–Adolescent version. *Eating and Weight Disorders: Studies on Anorexia, Bulimia and Obesity, 9,* 56–61.

Kozee, H. B., & Tylka, T. L. (2006). A test of objectification theory with lesbian women. *Psychology of Women Quarterly, 30,* 348–357.

Kutlesic, V., Williamson, D. A., Gleaves, D. H., Barbin, J. M., & Murphy-Eberenz, K. P. (1998). The Interview for the Diagnosis of Eating Disorders IV: Application to DSM-IV diagnostic criteria. *Psychological Assessment, 10,* 41–48.

Lau, A. S., Lum, S. K., Chronister, K. M., & Forrest, L. (2006). Asian American college women's body image: A pilot study. *Cultural Ethnicity and Minority Psychology, 12,* 259–274.

Levine, M. P., & Harrison, K. (2004). Media's role in the perpetuation and prevention of negative body image and disordered eating. In J. K. Thompson (Ed.), *Handbook of eating disorders* (pp. 695–717). New York: John Wiley.

Lindberg, S. M., Hyde, J. S., & McKinley, N. M. (2006). A measure of objectified body consciousness for preadolescent and adolescent youth. *Psychology of Women Quarterly, 30,* 65–76.

Lopez, E., Blix, G. G., & Blix, A. G. (1995). Body image of Latinas compared to body image of non-Latina White women. *Health Values: The Journal of Health, Behavior, Education, and Promotion, 19,* 3–10.

Lundgren, J., Anderson, D. A., & Thompson, J. K. (2004). Fear of negative appearance evaluation: Development and evaluation of a new construct for risk factor work in the field of eating disorders. *Eating Behaviors, 5,* 75–84.

Martins, Y., Tiggemann, M., & Kirkbride, A. (2007). Those Speedos become them: The role of self-objectification in gay and heterosexual men's body image. *Personality and Social Psychology Bulletin, 33,* 634–647.

Mayville, S., Katz, R. C., Gipson, M. T., & Cabral, K. (1999). Assessing the prevalence of body dysmorphic disorder in an ethnically diverse group of adolescents. *Journal of Child and Family Studies, 8,* 357–362.

Mayville, S. B., Williamson, D. A., White, M. A., Netemeyer, R., & Drab, D. L. (2002). Development of the muscle appearance satisfaction scale: A self-report measure for the assessment of muscle dysmorphia symptoms. *Assessment, 9,* 351–360.

McCreary, D. R., & Sasse, D. K. (2000). An exploration of the drive for muscularity in adolescent boys and girls. *Journal of American College Health, 48,* 297–304.

McCreary, D. R., Sasse, D. K., Saucier, D. M., & Dorsch, K. D. (2004). Measuring the drive for muscularity: Factorial validity of the drive for muscularity scale in men and women. *Psychology of Men & Masculinity, 5,* 49–58.

McKinley, N. M., & Hyde, J. S. (1996). The Objectified Body Consciousness Scale: Development and validation. *Psychology of Women Quarterly, 20,* 181–215.

Mendelson, B. K., Mendelson, M. J., & White, D. R. (2001). Body Esteem Scale for adolescents and adults. *Journal of Personality Assessment, 76,* 90–106.

Mendelson, B. K., & White, D. R. (1985). Development of self-body-esteem in overweight youngsters. *Developmental Psychology, 21,* 90–96.

Morrison, T. G., Morrison, M. A., Hopkins, C., & Rowan, E. T. (2004). Muscle mania: Development of a new scale examining the drive for muscularity in Canadian Males. *Psychology of Men and Masculinity, 5,* 30–39.

Noll, S. M., & Fredrickson, B. L. (1998). A mediational model linking self-objectification, body shame, and disordered eating. *Psychology of Women Quarterly, 22,* 623–636.

Olivardia, R. (2004). Body dysmorphic disorder. In J. K. Thompson (Ed.), *Handbook of eating disorders and obesity* (pp. 542–561). Hoboken, NJ: John Wiley.

Petersen, A. C., Schulenberg, J. E., Abramowitz, R. H., Offer, D., & Jarcho, H. D. (1984). A Self-Image Questionnaire for Young Adolescents (SIQYA): Reliability and validity studies. *Journal of Youth and Adolescence, 13,* 93–111.

Peterson, M., Ellenberg, D., & Crossan, S. (2003). Body-image perceptions: Reliability of a BMI-based silhouette matching test. *American Journal of Health Behavior, 27,* 355–363.

Petrie, T. A., Tripp, M. M., & Harvey, P. (2002). Factorial and construct validity of the Body Parts Satisfaction Scale–Revised: An examination of minority and non-minority women. *Psychology of Women Quarterly, 26,* 213–221.

Phillips, K. A. (1996). *The broken mirror.* New York: Oxford University Press.

Pook, M., Tuschen-Caffier, B., & Brähler, E. (2008). Evaluation and comparison of different versions of the Body Shape Questionnaire. *Psychiatry Research, 158,* 67–73.

Pope, H., Gruber, A., Choi, P., Olivardia, R., & Phillips, K. (1997). An underrecognized form of body dysmorphic disorder. *Psychosomatics, 38,* 548–557.

Pope, H., Katz, D., & Hudson, J. (1993). Anorexia nervosa and "reverse anorexia" among 108 male bodybuilders. *Comprehensive Psychiatry, 34,* 406–409.

Reas, D. L., Whisenhunt, B. L., & Netemeyer, R. (2002). Development of the Body Checking Questionnaire: A self-report measure of body checking behaviors. *International Journal of Eating Disorders, 31,* 324–333.

Reed, D. L., Thompson, J. K., Brannick, M. T., & Sacco, W. P. (1991). Development and validation of the Physical Appearance State and Trait Anxiety Scale (PASTAS). *Journal of Anxiety Disorders, 5,* 323–332.

Rguibi, M., & Belahsen, R. (2006). Body size preferences and sociocultural influences on attitudes towards obesity among Moroccan Sahraoui women. *Body Image, 3,* 395–400.

Rose, K. S., Cooper, M. J., & Turner, H. (2006). The eating disorder belief questionnaire: Psychometric properties in an adolescent sample. *Eating Behaviors, 7,* 410–418.

Rosen, J. C., & Reiter, J. (1996). Development of the Body Dysmorphic Disorder Examination. *Behaviour Research and Therapy, 34,* 755–766.

Rosen, J. C., Jones, A., Ramirez, E., & Waxman, S. (1996). Body shape questionnaire: Studies of validity and reliability. *International Journal of Eating Disorders, 20,* 315–319.

Rosen, J. C., Srebnik, D., Saltzberg, E., & Wendt, S. (1991). Development of a body image avoidance questionnaire. *Psychological Assessment, 3,* 32–37.

Roth, D., & Armstrong, J. (1993). Feelings of Fatness Questionnaire: A measure of the cross-situational

variability of body experience. *International Journal of Eating Disorders, 14,* 349–358.

Rousseau, A., Knotter, A., Barbe, P., Raich, R., & Chabrol, H. (2005). Validation of the French version of the Body Shape Questionnaire. *Encephale, 31,* 162–173.

Schulman, R. G., Kinder, B. N., Powers, P. S., Prange, M., & Gleghorn, A. A. (1986). The development of a scale to measure cognitive distortions in bulimia. *Journal of Personality Assessment, 50,* 630–639.

Shisslak, C. M., Renger, R., Sharpe, T., Crago, M., McKnight, K. M., Gray, N., et al. (1999). Development and evaluation of the McKnight Risk Factor Survey for assessing potential risk factor and protective factors for disordered eating in preadolescent and adolescent girls. *International Journal of Eating Disorders, 25,* 195–214.

Shroff, H., & Thompson, J. K. (2006a). Multiple aspects of influence for body image among adolescents: A test of the tripartite influence model. *Body Image, 3,* 17–23.

Shroff, H., & Thompson, J. K. (2006b). Peer influences, body image dissatisfaction, eating dysfunction and self-esteem in adolescent girls. *Journal of Health Psychology, 11,* 533–551.

Siegel, J. M., Yancey, A. K., Aneshensel, C. S., & Schuler, R. (1999). Body image, perceived pubertal timing, and adolescent mental health. *Journal of Adolescent Health, 25,* 155–165.

Spangler, D. L., & Stice, E. (2001). Validation of the Beliefs About Appearance Scale. *Cognitive Therapy and Research, 25,* 813–827.

Stein, K. E., & Hedger, K. M. (1997). Body weight and shape self-cognitions, emotional distress, and disordered eating in middle adolescent girls. *Archives of Psychiatric Nursing, 11,* 264–275.

Stice, E., Shaw, H., & Nemeroff, C. (1998). Dual pathway model of bulimia nervosa: Longitudinal support for dietary restraint and affect-regulation mechanisms. *Journal of Social and Clinical Psychology, 17,* 129–149.

Stunkard, A. J., Sorenson, T. I., & Schulsinger, F. (1983). Use of the Dutch Adoption Register for the study of obesity and thinness. In S. Kety, L. P. Rowland, R. L. Sidman, & S. W. Matthysse (Eds.), *The genetics of neurological and psychiatric disorders* (pp. 115–120). New York: Raven.

Szymanski, M. L., & Cash, T. F. (1995). Body-image disturbances and self-discrepancy theory: Expansion of the Body-Image Ideals Questionnaire. *Journal of Social and Clinical Psychology, 14,* 134–146.

Tantleff-Dunn, S., Thompson, J. K., & Dunn, M. E. (1995). The Feedback on Physical Appearance Scale (FOPAS): Questionnaire development and psychometric evaluation. *Eating Disorders: The Journal of Treatment and Prevention, 3,* 332–341.

Thelen, M. H., Powell, A. L., Lawrence, C., & Kuhnert, M. E. (1992). Eating and body image concern among children. *Journal of Clinical Child & Adolescent Psychology, 21,* 41–46.

Thompson, J. K. (2004). The (mis)measurement of body image: Ten strategies for improving assessment for clinical and research purposes. *Body Image: An International Journal of Research, 1,* 7–14.

Thompson, J. K., & Altabe, M. N. (1991). Psychometric qualities of the Figure Rating Scale. *International Journal of Eating Disorders, 10,* 615–619.

Thompson, J. K., Cattarin, J., Fowler, B., & Fisher, E. (1995). The Perception of Teasing Scale (POTS): A revision and extension of the Physical Appearance Related Teasing Scale (PARTS). *Journal of Personality Assessment, 65,* 146–157.

Thompson, J. K., & Gardner, R. (2002). Measuring perceptual body image among adolescents and adults. In T. F. Cash & T. Pruzinsky (Eds.), *Body image: A handbook of theory, research and clinical practice* (pp. 135–141). New York: Guilford.

Thompson, J. K., Heinberg, L. J., Altabe, M. N., & Tantleff-Dunn, S. (1999). *Exacting beauty: Theory, assessment and treatment of body image disturbance.* Washington, DC: American Psychological Association.

Thompson, J. K., Heinberg, L. J., & Tantleff, S. (1991). The Physical Appearance Comparison Scale. *The Behavior Therapist, 14,* 174.

Thompson, J. K., & Tantleff, S. T. (1992). Female and male ratings of upper torso: Actual, ideal, and stereotypical conceptions. *Journal of Social Behavior and Personality, 7,* 345–354.

Thompson, J. K., van den Berg, P., Roehrig, M., Guarda, A. S., & Heinberg, L. J. (2004). The Sociocultural Attitudes Towards Appearance Questionnaire-3 (SATAQ-3): Development and validation. *International Journal of Eating Disorders, 35,* 293–304.

Thompson, M. A., & Gray, J. J. (1995). Development and validation of a new body-image assessment tool. *Journal of Personality Assessment, 64,* 258–269.

Thurfjell, B., Edlund, B., Arinell, H., Hagglof, B., & Engstrom, I. (2003). Psychometric properties of Eating Disorder Inventory for Children (EDI-C) in Swedish girls with and without a known eating disorder. *Eating and Weight Disorders, 8,* 296–303.

Tiggeman, M., & Lynch, J. (2001). Body image across the life span in adult women: The role of self-objectification. *Developmental Psychology, 37,* 243–253.

Tiggemann, M., & Slater, A. (2001). A test of objectification theory in former dancers and non-dancers. *Psychology of Women Quarterly, 25,* 57–64.

Tylka, T. L., Bergeron, D., & Schwartz, J. P. (2005). Development and psychometric evaluation of the Male Body Attitudes Scale (MBAS). *Body Image, 2,* 161–175.

Veron-Guidry, S., & Williamson, D. A. (1996). Development of a body image assessment procedure for children and preadolescents. *International Journal of Eating Disorders, 20,* 287–293.

Wertheim, E. H., Paxton, S. J., & Tilgner, L. (2004). Test-retest reliability and construct validity of Contour Drawing Rating Scale scores in a sample of early adolescent girls. *Body Image, 1,* 199–205.

Williamson, D. A., Gleaves, D. H., Watkins, P. C., & Schlundt, D. G. (1993). Validation of the self-ideal body size discrepancy as a measure of body dissatisfaction. *Journal of Psychopathology and Behavioral Assessment, 15,* 57–68.

Wooley, O. W., & Roll, S. (1991). The Color-a-Person Body Dissatisfaction Test: Stability, internal consistency, validity, and factor structure. *Journal of Personality Assessment, 56,* 395–413.

Yamamiya, Y., Shroff, H., & Thompson, J. K. (2008). The tripartite influence model of body image and eating disturbance: A replication with a Japanese sample. *International Journal of Eating Disorders, 41,* 88–91.

Yanover, T., & Thompson, J. K. (2009) Assessment of body image in children and adolescents. In L. Smolak & J. K. Thompson (Eds.), *Body image, eating disorders and obesity in youth* (2nd ed., pp. 177–192). Washington, DC: American Psychological Association.

CHAPTER 5

Measures of Restrained Eating

Conceptual Evolution and Psychometric Update

Michael R. Lowe and J. Graham Thomas

Research on restrained eating was first published more than 30 years ago. During this period, much has been learned about the psychometric properties of the three primary measures of restrained eating. One purpose of the present chapter is to supplement and update the thorough review of restraint measures provided by Allison and Gorman in the first edition of this book. We have retained the psychometric information from their chapter and updated it with data published since 1995.

Since the first studies on restrained eating were published in 1975 (Herman & Mack, 1975; Herman & Polivy, 1975), another major development has been an evolution—some might say a revolution—in the very meaning of *restrained eating*. That is, at a conceptual level, it has become clearer what these measures are—and are not—measuring, and these newer findings are often inconsistent with the way in which the measures were originally conceptualized. Because the psychometric properties of a measure only become meaningful when some consensus exists on the concept or domain assessed by that measure, it is critical to review research that can help specify what *restrained eating* means and what restrained eating measures are assessing.

We first provide a brief historical overview of the three main measures of restrained eating and the rationale for their development. This is followed by a consideration of studies, most of which have been published since the first edition of this book, that have raised questions about what these scales are measuring. The final section provides an update on the psychometric properties of restraint scales, again focusing primarily on the three most widely used measures of restraint.

A Brief History of Measures of Restrained Eating

The first measure of restrained eating, the Restraint Scale (RS), was developed by Herman and Polivy (1975) with the final, 10-item revised

version published in 1978 (see Herman & Polivy, 1980). The rationale for the development of the RS grew out of the work of Schachter and Rodin (1974) and Nisbett (1972) on factors controlling food intake in the obese. Herman and Polivy (1975) and Herman and Mack (1975) reasoned that normal-weight individuals who were constantly dieting and holding their weight below its biological set point would demonstrate anomalies in eating behavior that resembled those shown by obese individuals in prior research. This reasoning directly followed from Nisbett's (1972) hypothesis that differences in eating behavior between normal-weight and overweight individuals were due to the overweight individuals keeping their body weight suppressed below its biologically appropriate level to conform to social norms. The RS was used to identify normal-weight individuals whose body weight was kept suppressed by constant dieting. Subsequent research has indeed found that the RS identifies normal-weight individuals who differ from unrestrained individuals on a wide variety of behavioral, cognitive, and physiological measures (Herman & Polivy, 1984; Lowe, 1993; Lowe & Kral, 2006).

Nonetheless, Stunkard and Messick (1985) noted several serious problems with the RS. First, they reviewed evidence showing that restrained, overweight individuals—unlike normal-weight restrained eaters—did not overeat following consumption of a preload. This contradicted Herman and Polivy's (1975) assumption that previously observed differences in eating behavior between normal-weight and obese individuals were due to greater dieting in the obese. If this were true, then restrained obese individuals should be at least as susceptible to preload-induced overeating as normal-weight restrained eaters. Second, they noted that the RS measured not only dietary restraint but also weight fluctuations. Weight fluctuation is often higher in obese individuals for reasons having nothing to do with dieting behavior (e.g., if the degree of weight fluctuation that individuals experience is a constant fraction of their body mass, obese individuals will experience larger fluctuations in absolute terms; see Drewnowski, Riskey, & Desor, 1982). Furthermore, Drewnowski et al. (1982) showed that two weight fluctuation items account for 70% of the variance in total RS scores and also found that obese persons actually scored lower on the dietary concern factor of the RS. These problems led Stunkard and Messick to develop a new measure of restraint—the restrained eating scale—which is one of three factors in their Three-Factor Eating Questionnaire (TFEQ; now called the Eating Inventory [EI]). The TFEQ-R scale represented a major improvement in the assessment of restrained eating because it eliminated two confounds—between dieting and overeating and between restrained eating and overweight—that characterize the RS. Investigators studying restrained eating broadly agree that the TFEQ-R, relative to the RS, represents a "purer" measure of restraint that is more likely to reflect actual efforts to restrict dietary intake (Stunkard & Messick, 1985; van Strien, 1999).

van Strien, Frijters, Bergers, and DeFares (1986) noted many of the same problems with the RS that Stunkard and Messick (1985) described. To address these limitations, they developed the Dutch Eating Behavior Questionnaire (DEBQ) that included a restrained eating scale. Their restraint scale is quite similar to Stunkard and Messick's TFEQ-R measure, in part because both groups used items from a measure developed by Pudel, Metzdorff, and Oetting (1975) to construct their scales. Pudel et al.'s scale assessed "latent obesity" or the tendency of some normal-weight individuals to exhibit eating patterns previously associated with obesity. A major advantage of both the TFEQ-R and DEBQ-R scales is that they (unlike the RS) reflect "pure" dietary restraint, permitting the dissociation of restrained eating from its opposite—overconsumption. Interestingly, the TFEQ-R and DEBQ-R scales have weak or nonexistent relationships with the other subscales of the Eating Inventory and DEBQ that tap different types of excessive eating (stemming from disinhibition, hunger, negative emotions, and external food stimuli).

Developments in the Definition and Conceptualization of Restrained Eating

The majority of early theorizing about restrained eating and its possible psychobiological effects was produced by Herman, Polivy, and their colleagues (Herman & Polivy, 1984). However, during the past 30 or so years, there have been a variety of research findings that have (a) raised questions about the meaning and definition of restrained eating and (b) shed new light on what measures of restrained eating are—and are not—assessing. In the next several sections, we review this literature because it suggests that traditional theorizing about restrained eating and its putative effects is in need of major revision.

The Meaning of Restrained Eating

Most researchers have defined restrained eating in a manner similar to the following: Restrained eating refers to conscious efforts to restrict food intake for the purpose of weight control. Over the years, Herman and Polivy have defined the term in different ways, including the suppression of body weight below one's body weight set point (Herman & Polivy, 1975), the imposition of a cognitively defined "diet boundary" to limit food intake (Herman & Polivy, 1984), and a history of repeatedly going on and off diets, referred to as "unsuccessful dieting" (Heatherton, Herman, Polivy, King, & McGree, 1988).

From one perspective, it is certainly possible that a restrained eater on the RS would have all three characteristics implied by these definitions: that her weight would be well below her highest weight ever (and perhaps therefore well below her body weight "set point"), that she would impose a diet boundary on her eating to establish permissible caloric intake, and that she had been on and off diets repeatedly in the past. However, as Lowe (1993) pointed out, these are characteristics that could also be used to differentiate three different *types* of restrained eaters: those who (a) are well below their highest weights ever by virtue of intentional weight loss (weight suppressors), (b) are currently on a diet to lose weight (current dieters), or (c) have engaged in repeated cycles of dieting and overeating in the past (frequent dieters and overeaters). Importantly, even though most restrained eaters will be characterized by one or more of these designations, these three types of dieting are theoretically independent: Knowing a person's status on any one of these three dieting types does not necessarily tell you anything about his or her standing on the others. Furthermore, because these different dieting patterns are associated with different appetitive and behavioral responses, Lowe pointed out that measuring a single construct of "restraint" could conceal important differences between these dieting subtypes. For instance, Lowe, Whitlow, and Bellwoar (1991) found that restrained eaters who were not currently dieting ate somewhat more with than without a preload, whereas current dieters ate much less with than without a preload. Another example comes from Lowe, Thomas, Safer, and Butryn (2007), who recently reported that weight suppression was positively associated with binge-eating frequency among individuals diagnosed with bulimia nervosa (which is consistent with Russell's [1979] original theorizing about the role of significant weight losses in bulimia), whereas scores on the Eating Disorders Examination–Restrained Eating subscale were negatively associated with binge-eating frequency (which is inconsistent with the cognitive-behavioral model of bulimia).

Another major development regarding the meaning of restraint involves the motivation underlying restrained eating. At different times, Herman and Polivy have viewed restrained eating as a way of examining the effects of dieting to avoid weight gain (Herman & Polivy, 1975), as a factor contributing to weight gain and obesity (Polivy & Herman, 1983), and as a major cause of eating disorders (Polivy & Herman, 1985). Despite these wide variations in the purported significance of restrained eating, the same scale (the RS) has been used to measure restraint, and

essentially the same theory (that dieting behavior sows the seeds of its own destruction) has been used to account for these varied functions of restrained eating. However, when particular subtypes of dieting are examined in relation to outcomes traditionally studied in the restraint literature, the need to go beyond the use of a single, monolithic measure of restraint emerges again. For example, weight suppressors, who, according to the set point model (Herman & Polivy, 1975), should be hyperresponsive to appetitive stimuli, instead show rigorous eating control following a preload (Lowe & Kleifield, 1988) and reduced sweetness preferences (Kleifield & Lowe, 1991). When applied to those with bulimia nervosa, restraint theory predicts that bulimic individuals who are currently dieting should binge more than bulimic nondieters; instead, the opposite relationship has been found in two studies (Lowe, Gleaves, & Murphy-Eberenz, 1998; Lowe et al., 2007).

These findings indicate that the advisability of using the RS to study restrained eating depends on the investigator's research objectives. If the objective is to study "restrained eating" as operationalized by the RS (which involves the simultaneous measurement of several constructs—dieting, overeating, weight fluctuations, overweight), then the RS could be appropriate to use—and has the advantage of being characterized by a large corpus of previously published findings. (Alternatively, investigators are increasingly using a combination of the TFEQ-R and disinhibition scales from the EI [e.g., Westenhoefer, Broeckmann, Munch, & Pudel, 1994; Williamson et al., 1995]—based on the reasonable assumption that those who score high on both resemble restrained eaters measured by the RS—because this approach permits them to study both the independent and interactional relationship of restrained eating and predisposition toward overeating on outcomes of interest.) If the objective of a study is to examine particular types of dieting, on the other hand, then an alternative to the RS should be considered—for example, by measuring one of the three types of dieting behavior outlined by Lowe (1993) or by putting people on short-term weight loss diets (Presnell & Stice, 2003).

What Are Restraint Scales Measuring?

As noted above, Herman and Polivy (1975) originally conceived of restrained eaters as individuals who were "constantly dieting and concerned with not gaining weight, and who presumably would gain substantial weight if they were to 'let themselves go'" (p. 667). Although these authors subsequently de-emphasized this characterization of restrained eaters in favor of a more cognitively focused perspective (Herman & Polivy, 1984) that emphasized drive for thinness (Polivy & Herman, 1987), it appears—as we shall see shortly—that this original viewpoint may actually best capture the nature of restrained eaters' vulnerability to aberrations in their appetitive and consummatory responses.

Twenty-five years ago, when no one realized that developed countries were entering the early stages of explosive growth in the prevalence of obesity, dieting in normal-weight individuals (and normal-weight women in particular) was assumed to reflect an unhealthy need to achieve a slim body to conform with societal norms of attractiveness (Striegel-Moore, Silberstein, & Rodin, 1986). As Polivy and Herman put it in 1987, "Nowadays, women are induced to strive toward a condition of ruddy-cheeked emaciation" (p. 635). This emphasis on attaining the "thin ideal" has been widely accepted as the primary driver of restrained eating among individuals in the normal weight range. Thus, restraint theory has gone 180 degrees from its original belief that restrained eating is motivated by an effort to prevent weight gain (Herman & Polivy, 1975) to the belief that it is motivated by the yearning for an unrealistically thin body (Polivy & Herman, 1987).

These seemingly contradictory possibilities might be clarified by drawing two distinctions regarding restrained eaters' motivation for weight control. The first distinction involves restrained

eaters' goals for weight change. Restrained eaters have elevated levels of body dissatisfaction (Ruderman & Grace, 1988) and both desire a thinner body (Polivy & Herman, 1987) and fear weight gain (Vartanian, Herman, & Polivy, 2005). Presumably, most restrained eaters would like to consume fewer calories than they expend and lose some weight, thereby moving closer to their desired goal and further away from the feared outcome of weight gain. However, Polivy and Herman's assumption that restrained eaters are driven to reach unhealthy levels of body weight conflates restrained eaters' desire to be *thinner* (e.g., to lose a few pounds) with their desire to be objectively *thin* (e.g., to achieve a body weight far below their medically appropriate weight for their height). A recent study (Chernyak & Lowe, 2007) compared unrestrained and restrained eaters on drive for thinness, fear of fatness, and drive to be objectively thin (defined as being 15% below their medically appropriate weight for height). Restrained eaters scored significantly higher than unrestrained eaters on the first two measures but did not differ from unrestrained eaters on the third measure. These findings suggest that while restrained eaters would like to avoid weight gain or to lose a small amount of weight, they do not have an unhealthy drive to become pathologically thin. Therefore, it appears that restrained eaters are not as strongly motivated to lose weight as has often been assumed. The fact that most restrained eaters are not currently dieting to lose weight (Lowe, 1993) is consistent with this conclusion.

The second distinction involves the extent to which restrained eaters, whatever their weight control goals, are actually reducing their caloric intake. Restrained eaters assessed with the RS are assumed to vacillate between periods of caloric restriction and overindulgence without losing weight in absolute terms (Heatherton, Polivy, & Herman, 1991). Restrained eaters on the other two restraint measures are generally assumed to be more successful at caloric restriction, especially since these measures are viewed as purer measures of the actual cognitions and behaviors involved in dieting (Stunkard & Messick, 1985; van Strien, 1999). However, although past lab studies sometimes found that restrained eaters consume less food than unrestrained eaters, recent evidence indicates that restrained eaters, no matter how they are identified, do not eat less in the natural environment than unrestrained eaters (Stice, Cooper, Schoeller, Tappe, & Lowe, 2007).

Stice, Fisher, and Lowe (2004) examined five dietary restraint scales that were developed to assess intentional dietary restriction for the purposes of weight control. These scales showed weak and generally nonsignificant correlations with objectively measured caloric intake during unobtrusively observed eating episodes across four studies (mean $r = -.07$; range: $-.34$ to .20; Stice et al., 2004). For example, the average correlation between three dietary restraint scales and observed caloric intake of students consuming meals in a cafeteria was $-.09$.

In response to these validity findings, van Strien, Engels, van Staveren, and Herman (2006) noted that short-term caloric intake may not be representative of long-term caloric intake and suggested that researchers test whether dietary restraint scales show inverse correlations with objective measures of longer term caloric intake. Four previous studies (reviewed in Stice et al., 2007) that examined this question found no relationship between caloric intake and several measures of restrained eating. In a recent follow-up study, Stice et al. (in 2007) reported on three additional studies that found that the TFEQ-R scale was not correlated with doubly labeled water-estimated energy intake over 2-week periods or with observationally measured caloric intake over 3 months. Taken together, the foregoing findings suggest that dietary restraint scales may not be valid measures of naturalistic dietary restriction and imply the need to reinterpret findings from studies that have used dietary restraint scales. As Lowe and Levine (2005), Lowe and Butryn (2007), and Stice et al. (2007) have recently suggested, part of this reinterpretation should be based on the idea that measures of restrained eating reflect *relative* dietary restriction (i.e., relative to the positive energy balance

that would result if a restrained eater no longer practiced restraint) rather than *absolute* dietary restriction (i.e., relative to energy balance or to the intake of unrestrained eaters).

Two other teams of researchers have come to similar conclusions. First, Gorman, Allison, and Primavera (1993) and Allison, Kalinsky, and Gorman (1992) conducted a factor analysis of the TFEQ-R scale and found that it contained two factors that they called cognitive restraint and behavior restraint. They analyzed their data using nonlinear techniques that take into account situations where items differ substantially in their endorsement rates. Their results suggested that the TFEQ-R items form a continuum that begins with relatively common thoughts of reducing eating and ends with overt, deliberate, but relatively rare actions to reduce eating. These results indicate that even measures that ostensibly reflect "successful" restraint do not identify individuals who eat less than unrestrained eaters.

Second, Larsen, van Strien, Eisinga, Herman, and Engels (2007) recently factor analyzed the DEBQ-R among a large sample of weight-concerned individuals and found that a two-factor solution fit the data well. The two factors differentiated between restrained eating intentions and restrained eating behavior. In line with Allison et al.'s (1992) work, they found that participants scored higher on dieting intentions than dieting behavior. They also found that that more restrained eating behavior was related to "*less* external and emotional eating, whereas more restrained intentions (without restrained behavior) were related to *more* external and emotional eating" (p. 106). These results are reminiscent of the distinction Lowe et al. (1991) made between restrained eaters who are and are not currently dieting to lose weight, with the former group showing a counterregulatory eating pattern and the latter group showing eating regulation.

We should also note that the same questions about the relationships between food restriction, overeating, and weight control have been raised in research in children. Birch, Fisher, and Davison (2003) found that 5-year-old girls whose mothers reported using restrictive feeding practices were more likely to exhibit eating in the absence of hunger at 9 years of age. This was especially true of girls who were already overweight at the age of 5. These results are suggestive of a gene-by-environment interaction in which overweight girls are genetically predisposed to be highly sensitive to environmental influences over eating. There is no way of knowing from these data whether mothers' restrictive eating practices are causally related to later vulnerability to eating in the absence of hunger, if they reflect mothers' concerns about concurrent weight gain in their children and have no causal influence, or if this relationship is due to some as yet unidentified variable.

All the research reviewed in this section suggests that, despite their *desire* to be thinner, in functional terms, most restrained eaters are at best employing restraint to avoid weight gain, not to lose weight. This conclusion is supported by research showing that measures of restrained eating prospectively predict weight gain rather than weight loss (French, Jeffery, & Wing, 1994; Klesges, Isbell, & Klesges, 1992; Stice, Presnell, Shaw, & Rohde, 2005). It appears that, just as most obese individuals who lose weight via dieting eventually regain it (Sarwer & Wadden, 1999), restrained eating may forestall but usually does not prevent weight gain. One additional reason to suggest that much of restrained eaters' motivation for weight control stems from concerns about gaining weight is that restrained eaters show levels of certain hormones (e.g., reduced leptin, increased cephalic phase insulin) that makes them metabolically predisposed toward weight gain (Lowe & Kral, 2006). Although these findings theoretically could be due to metabolic adaptations to weight loss dieting, the evidence reviewed above indicates that restrained eaters are not in negative energy balance. In sum, it appears that our understanding of the nature of the motivation that has fueled the tremendous increase in dieting behavior in the past few decades has come full circle. Herman and Polivy (1975) started out believing that restrained eating was driven by the desire to avoid weight gain secondary to being

below one's biologically determined body weight set point value. If one replaces the notion that a body weight set point is "pulling" weight upward from within (e.g., via the hypothalamus [Nisbett, 1972]) with the idea that an obesogenic environment is "pulling" weight upward from without (Lowe & Butryn, 2007; Lowe & Levine, 2005) then Herman and Polivy's original theorizing appears to be closest to the truth. That is, normal-weight restrained eaters and dieters appear to have a predisposition toward weight gain in an obesogenic environment (e.g., Lowe et al., 2006), but this characteristic would presumably remain latent in environments where food was difficult to come by. From this perspective, the fact that the first research on restrained eating was conducted around the same time that the obesity epidemic began is probably no coincidence.

One caveat is needed before bringing this section of the chapter to a close. The fact that measures of restrained eating generally do not reflect caloric restriction or weight loss dieting should not be taken to mean that diet-induced weight loss is not a risk factor for the development of eating disorders. On one hand, it does appear that the multiple findings in the literature showing that measures of restrained eating or dieting prospectively predict increased bulimic symptoms are not due to low-calorie dieting (Stice et al., 2007). On the other hand, there is good evidence that extreme dieting that produces rapid, extensive weight loss may indeed help cause bulimia nervosa (Butryn & Wadden, 2005; Garner & Fairburn, 1988; Keys, Brozek, Henschel, Mickelsen, & Taylor, 1950; Russell, 1979). These findings are a further indication that it behooves researchers to think carefully about precisely what construct they are interested in investigating when studying "restrained eating" and to tailor their measures of that construct accordingly.

Herman and Polivy's Restraint Scale

Description

Herman and Mack (1975) originally developed the Restraint Scale (RS) to identify normal-weight individuals who attempt to limit their food intake in an effort to resist biological pressures toward weight gain. The original scale consisted of 5 items measuring chronic dieting. The items were rationally derived and selected for face validity. The scale was tested on a sample of 45 women, which produced a Cronbach's alpha coefficient of 0.65. Herman and Polivy (1975) revised the instrument to include 11 items, with 6 items forming a Diet and Weight History subscale (alpha coefficient 0.62) and the remaining 5 items forming a Concern With Dieting subscale (alpha coefficient 0.68). The subscales correlated at 0.48, and the alpha coefficient for the whole scale was 0.75. The final iteration of the RS (Herman & Polivy, 1980) consists of 10 items. Polivy, Herman, and Howard (1988) describe the RS as "a 10-item self report questionnaire assessing weight fluctuations, degree of chronic dieting, and related attitudes toward weight and eating" (p. 377). The preponderance of published research using the RS has used this 10-item version.

Herman and Polivy (1975) subdivided the RS into two subscales. The Weight Fluctuation (WF) subscale (Items 2, 3, 4, and 10) measures both instability in weight and a history of overweight. The Concern for Dieting (CD) subscale (Items 1, 5, 6, 7, 8, and 9) assess preoccupation with food, overconcern about eating, and overeating tendencies. Thus, an individual who scores highly on both subscales is likely to be characterized by a history of overweight, a desire to weigh less, and unstable body weight. Notably, the RS should not be considered a measure

of actual hypocaloric dieting or energy deficit (Polivy et al., 1988). High scores on the RS are prospectively associated with greater fluctuations in body weight (Heatherton, Polivy, & Herman, 1991; Tiggemann, 1994). Some (Heatherton et al., 1988) have cited the link between restraint scores and weight fluctuation as support for the idea that the concept of restraint should include efforts to restrict eating to control one's weight *and* the periodic failure of restraint resulting in episodes of overeating (i.e., disinhibited eating). The RS is consistent with this formulation of restraint and the associated theory that dieting is a major cause of overeating and eating disorders (Polivy & Herman, 1985). As noted previously, this theory has undergone increased scrutiny (e.g., Lowe & Kral, 2006; Stice et al., 2007).

Sample

The RS was initially tested on samples of 42 (Herman & Mack, 1975) and 45 (Herman & Polivy, 1975) female college students. The great majority of psychometric studies using the RS have been done with normal-weight and overweight female college students. It is occasionally used with eating-disordered individuals but rarely with clinical populations of overweight individuals.

Norms

Studies that incorporate the RS as a measure of primary interest tend to use the RS in one of two ways: either as a continuous measure of restrained eating or as a tool to dichotomize a sample into restrained and unrestrained eaters. In the former case, the RS is typically analyzed with regression methods to investigate constructs that may be associated with restraint. This analytic strategy is desirable because it preserves the full variability of the RS. In the latter case, after participants have been classified as restrained eaters or unrestrained eaters, the two groups are compared on some measure(s), often in an analysis of variance (ANOVA). Historically, the latter approach was more common than the former. Typically, a median split was used to create groups of restrained and unrestrained eaters of approximately equal size. However, medians varied across samples, which resulted in different cutoffs for identifying restrained eaters. Concern over the failure to consistently identify a homogeneous set of restrained eaters across studies led some researchers to adopt the most frequently observed RS median (a score 15) as the standard cutoff for use in studies of restrained eating. This approach has the strength of standardization of the definition of restrained eaters but also the weakness of using dissimilar strategies of defining restrained eating in earlier and later studies of this construct. Furthermore, there is some evidence that medians on the RS are decreasing over time (e.g., medians were in the 15–17 range in the 1970s but are most often in the 12–14 range more recently), which casts doubt on the utility of a preselected cutoff to identify restrained eaters.

Table 5.1 presents sample sizes, means, and standard deviations for the RS, as well as its subscales, for a variety of samples. The average score for normal-weight women (mostly from samples of college students) is about 13. The corresponding value for men is 10. These values are useful for determining whether a particular sample is unusually high or low on restraint. It is important to keep in mind that RS scores may differ by nationality, weight status, eating disorder status, or other personal characteristics.

Table 5.1 Mean Restraint Scale Scores Reported in the Literature

Scale and Participants	*Author*	*n*	*Mean*	*SD*
Whole scale				
American adults Female Male	French, Jeffery, and Wing (1994)	 103 99	 14.6 11.0	 5.5 5.0
American adult women	Timmerman and Gregg (2003)	120	20.5	4.6
American overweight adults	Williamson et al. (2007)	46	13.4	6.0
American college students Female Male	Allison, Kalinsky, and Gorman (1992)	901 617 282	15.1 16.4 12.3	7.0 6.9 6.4
American college students Female Male	Boerner, Spillane, Anderson, and Smith (2004)	 215 214	 13.0 8.9	 6.1 5.5
American college students Female Male	Klem, Klesges, Bene, and Mellon (1990)	497 346 151	12.6 13.4 10.8	5.9 5.8 5.8
American female college students	Urland and Ito (2005)	82	13.8	6.6
Australian female college students	Griffiths et al. (2000)	82	12.1	6.0
British adolescent women	Cole and Edelmann (1987)	184	10.6	5.9
British women	Wardle and Beales (1987)	102	13.5	5.4
British men	Wardle and Beales (1987)	45	8.5	5.8
Canadian college students Female Male	Oates-Johnson and DeCourville (1999)	220 159 61	11.6 12.8 8.5	6.6 6.5 5.7
Canadian college students Female Male	Rotenberg and Flood (2000)	 159 61	 12.8 8.5	 6.5 5.7
Dutch obese women	Westerterp-Plantenga, Kempen, and Saris (1998)	57	20	3.5
Portuguese female college students	Scagliusi et al. (2005)	62	11.3	5.0

(Continued)

Table 5.1 (Continued)

Scale and Participants	*Author*	*n*	*Mean*	*SD*
Portuguese women				
With anorexia nervosa	Scagliusi et al. (2005)	15	17.3	9
With bulimia nervosa	Scagliusi et al. (2005)	24	28.1	13.0
Weight fluctuation scale				
Overweight American adults	Williamson et al. (2007)	46	6.9	3.2
American college students	Allison et al. (1992)	901	5.8	3.3
Female		617	5.9	3.3
Male		282	5.3	3.5
American college students	Boerner et al. (2004)			
Female		215	5.3	2.9
Male		214	4.8	3.3
American college students	Klem, Klesges, Bene, et al. (1990)			
Female		346	5.0	
Male		151	4.8	
British adolescent women	Cole and Edelmann (1987)	184	4.1	2.8
British women	Wardle and Beales (1987)	102	4.9	2.8
British men	Wardle and Beales (1987)	45	3.1	3.1
Concern with dieting scale				
American adults	Williamson et al. (2007)	46	6.4	4.7
American college students	Allison et al. (1992)	901	9.3	4.7
Female		617	10.4	4.6
Male		282	6.9	3.9
American college students	Boerner et al. (2004)			
Female		215	7.7	3.9
Male		214	4.0	3.2
American college students	Klem, Klesges, Bene, et al. (1990)			
Female		346	8.4	4.0
Male		151	5.9	3.4
British adolescent women	Cole and Edelmann (1987)	184	6.2	3.5
British women	Wardle and Beales (1987)	102	7.8	3.7
British men	Wardle and Beales (1987)	45	4.7	3.0

It appears that the practice of dichotomizing individuals into restrained and unrestrained eaters as the primary method of analysis should be discontinued. Stein (1988) demonstrated that an ANOVA design in which participants are dichotomized into groups based on RS score may have less predictive power in a preload study than a regression model in which RS scores are treated as continuous. Maxwell and Delaney (1993) confirmed that using median splits to form factors in a grouped design reduces statistical power. The authors also reported that dichotomizing participants based on a median split may produce erroneous conclusions about interactions among factors. This is especially relevant, as interactions between restraint status and various disinhibiting stimuli form the basis of many studies on restrained eating. Allison, Gorman, and Primavera (1993) discussed the disadvantages of dichotomization in general. Given these findings, the use of full RS scores in regression models is encouraged over the dichotomization of participants into restraint groups based on a median split. An exception may be made in situations where a strong theoretical or empirical basis exists for identifying specific groups of participants based on their RS score. Furthermore, though treating restraint scores continuously is preferable, because the majority of studies have treated restraint as a dichotomy, it is a good idea for researchers to analyze their results both ways so the categorical results can more easily be compared with past studies. A final reason to analyze results dichotomously in secondary analyses is that certain outcomes showing a nonlinear distribution may produce significant results using a median split but not with using continuous scores.

Age

There are a dearth of studies examining the relationship between the RS and age. Two studies including college students ranging from ages 17 to 57 years failed to find a relationship between RS and age (Allison et al., 1992; Klem, Klesges, Bene, & Mellon, 1990). However, little can be concluded from these studies because the most participants were between the ages of 18 and 22 years. The point in the human life span when dietary restraint typically asserts itself is unknown. A study by Cole and Edelmann (1987) observed a typical distribution of restraint scores in a sample of adolescent women with a mean age of 15 years old (see Table 5.1).

Gender

Boerner, Spillane, Anderson, and Smith (2004) observed higher RS total scores and CD subscale scores, but not WF subscale scores, among college women as compared with men. This pattern was also found by Allison et al. (1992). Klem, Klesges, Bene, and Mellon (1990) found that college women scored higher than men on the CD subscale but not the WF subscale or total RS. Oates-Johnson and DeCourville (1999) reported that college women scored significantly higher than men on the RS. This pattern was also observed by Rotenberg and Flood (2000). French et al. (1994) found that women scored higher on the RS than men in a sample of 202 adults, about a quarter of whom reported that they were actively dieting to lose or maintain weight. In a sample of adults, Klesges et al. (1992) reported higher restraint score in women than men. Thus, women appear to report systematically greater restraint on the RS than men. Concern for dieting seems to be more responsible for this difference than a history of weight fluctuation.

Reliability

Internal Consistency

The RS has been shown to have good internal consistency (Cronbach's alpha greater than .75) when used with normal-weight, non-eating-disordered samples. Table 5.2 illustrates the lower alpha levels that are observed in overweight and eating-disordered groups. This difference in alphas is likely attributable to restricted range within the overweight and eating-disordered subgroups. Crocker and Algina (1986) point out that Pearson product-moment correlations are lower when the variance of one or more variables in the analysis is restricted. As alpha depends on both the number of items and the correlation between items, subgroups of participants who respond in a systematically similar manner will produce lower alphas than more diverse subgroups that show greater variability in their scores. Drewnowski et al. (1982) were the first to point out that overweight and obese individuals are likely to score highly on the RS, and specifically the WF subscale, as a result of large weight fluctuations due, at least in part, to their increased adipose tissue rather than to restrained eating or concern about their weight. This potential measurement artifact may be a further source of homogeneity and subsequent lower internal consistency among overweight samples.

The CD and WF subscales show predictably lower alpha levels than the full RS score, presumably due to their smaller number of items. The alphas range from .66 to .71 for the CD subscale and from .70 to .80 for the WF subscale (Allison et al., 1992; Herman & Polivy, 1975; Klem, Klesges, Bene, & Mellon, 1990). van Strien, Breteler, and Ouwens (2002) and van Strien, Herman, Engels, Larsen, and van Leeuwe (2007) examined the internal consistency of the CD and WF subscales after removing Item 6 because of possible criterion confounding (Stice, Ozer, & Kees, 1997) and Item 10 because of inconsistent subscale factor loadings (Blanchard & Frost, 1983; Lowe, 1984; Overduin & Jansen, 1996). The resulting alphas for a group of 209 Dutch female college students were .77 for the five CD items and .70 for the three WF items (van Strien et al., 2002). van Strien et al. (2007) replicated the analysis with 349 normal-weight Dutch female college students and 409 overweight Dutch women and found alphas of .81 and .68 for the altered CD and WF scales for the normal-weight students and alphas of .65 and .72 for the altered CD and WF scales for the overweight women. Boerner et al. (2004) found that alphas for the total RS and its subscales are slightly higher for women than men. Klem, Klesges, Bene, et and Mellon (1990) determined that alphas for the RS and its subscales are equivalent for men and women, as well as for Blacks and Whites.

Test-Retest Reliability

RS scores appear to be stable over time (see Table 5.3). A somewhat lower coefficient was obtained with the Scagliusi et al. (2005) Portuguese translation of the RS.

Validity

There are a multitude of studies linking the RS to various aspects of eating behaviors, psychopathology, personality factors, and other constructs. It is beyond the scope of this chapter to review all of these reports; furthermore, most of them were not designed to test the validity of the RS. Rather, studies have been included that were (a) designed explicitly to test the

Table 5.2 Reliability of Dietary Restraint Scales: Internal Consistency

Reference	*n*	*Coefficient alpha*	*Sample Characteristics*
RS			
Allison, Kalinsky, and Gorman (1992)	823	.83	Normal-weight college students
Allison et al. (1992)	78	.72	Obese college students
Allison et al. (1992)	901	.82	Above two samples combined
Boerner, Spillane, Anderson, and Smith. (2004)	214	.76	Male college students
Boerner et al. (2004)	215	.82	Female college students
Laessle, Tuschl, Kotthaus, and Pirke (1989)	60	.78	Normal-weight women 18 to 30 years old; mostly college students
Rudderman (1983)	89	.86	Normal-weight female college students
Rudderman (1983)	58	.51	Obese female college students
W. G. Johnson, Lake, and Mahan (1983)	51	.79	Normal weight
W. G. Johnson et al. (1983)	58	.50	Obese nondieters
W. G. Johnson et al. (1983)	27	.83	Obese dieters
W. G. Johnson et al. (1983)	26	.57	Bulimic women 13 to 41 years old
Klem, Klesges, Bene, and Mellon (1990)	497	.78	College students (151 men; 346 women)
Klem, Klesges, Bene, et al. (1990)	124	.68	Obese college students
Klem, Klesges, Bene, et al. (1990)	373	.78	Normal weight college students
Oates-Johnson and DeCourville (1999)	220	.84	College students (61 men; 159 women)
Ouwens, van Strien, van der Staak (2003)	209	.83	Female college students
Rotenberg and Flood (1999)	58	.78	Female college students
Rotenberg and Flood (2000)	319	.77	College students (112 men; 207 women)
Urland and Ito (2005)	82	.85	Normal-weight female college students
van Strien, Cleven, and Schippers (2000)	200	.73	Female college students
van Strien, Herman, Engels, Larsen, and van Leeuwe (2007)	349	.84	Normal-weight female college students
van Strien et al. (2007)	409	.73	Overweight, nonobese women

(Continued)

Table 5.2 (Continued)

Reference	*n*	*Coefficient alpha*	*Sample Characteristics*
TFEQ-R			
Allison et al. (1992)	823	.91	Normal-weight college students
Allison et al. (1992)	78	.88	Obese college students
Allison et al. (1992)	901	.90	Above two samples combined
Boerner et al. (2004)	214	.89	Male college students
Boerner et al. (2004)	215	.90	Female college students
Laessle et al. (1989)	60	.80	Normal-weight women 18 to 30 years old; mostly college students
Ouwens et al. (2003)	209	.88	Female college students
Ricciardelli, Tate, and Williams (1997)	171	.91	Female college students
Simmons, Smith, and Hill (2002)	392	.87	American female 7th graders
Simmons et al. (2002)	300	.88	American female 10th graders
Stunkard and Messick (1988)	45	.92	Unrestrained eaters
Stunkard and Messick (1988)	53	.79	Restrained eaters
Stunkard and Messick (1988)	98	.93	Above two samples combined
van Strien et al. (2000)	200	.80	Female college students
DEBQ-R			
Allison et al. (1992)	823	.95	Normal-weight college students
Allison et al. (1992)	78	.91	Obese college students
Allison et al. (1992)	901	.95	Above two samples combined
Banasiak et al. (2001)	393	.94	Grade 9 female adolescents
Laessle et al. (1989)	60	.89	Normal-weight women 18 to 30 years old; mostly college students
Ouwens et al. (2003)	209	.94	Female college students
van Strien, Frijters, Bergers, and Defares (1986)	114	.94	Obese adults (71 men; 73 women)
van Strien, Frijters, et al. (1986)	996	.95	Normal-weight adults (427 men; 569 women)
van Strien, Frijters, et al. (1986)	1169	.95	Above two samples combined
van Strien et al. (2000)	200	.94	Female college students
van Strien et al. (2007)	349	.93	Normal-weight female college students
van Strien et al. (2007)	409	.89	Overweight, nonobese women

NOTE: DEBQ-R = Dutch Eating Behavior Questionnaire–Restraint subscale; RS = Restraint Scale; TFEQ-R = Three-Factor Eating Questionnaire–Restraint scale.

Table 5.3 Reliability of Dietary Restraint Scales: Test-Retest Reliability

Reference	*n*	*Coefficient*	*Interval*	*Sample Characteristics*
RS				
Allison, Kalinsky, and Gorman (1992)	34	.95	2 weeks	College students
Hibscher and Herman (1997)	86	.92	"A few weeks"	Male college students
Polivy, Herman, and Howard (1988)	514	.93	1 week	College students (166 men; 348 women)
Kickham and Gayton (1977)	44	.93	4 weeks	Normal-weight college students (16 men; 28 women)
Klesges, Klem, Epkins, and Klesges (1991)	305	.74	2½ years	98 men, 207 women
Scagliusi et al. (2005)	50	.64	1 month	Female college students
TFEQ-R				
Allison et al. (1992)	34	.91	2 weeks	College students
Bond, McDowell, and Wilkinson (2001)	64	.81	1 year	College students
Stunkard and Messick (1988)	17	.93	4 weeks	College students
DEBQ-R				
Allison et al. (1992)	34	.92	2 weeks	College students
Banasiak et al. (2001)	165	.85	4 to 5 weeks	High school students

NOTE: DEBQ-R = Dutch Eating Behavior Questionnaire–Restraint subscale; RS = Restraint Scale; TFEQ-R = Three-Factor Eating Questionnaire–Restraint scale.

validity of the RS or (b) report results that may be interpreted after the fact as support for, or evidence against, the theoretical assumptions that serve as a foundation for the development and continued use of the RS.

Content Validity

The RS was originally designed to be used with normal-weight individuals. Furthermore, factor-analytic studies of the Restraint Scale often obtain different factor solutions as a function of the number of overweight participants in the sample. Thus, studies including a large proportion of overweight participants are covered in their own section below.

Factorial Composition in Primarily Normal-Weight Samples

The two-factor model of the RS, including the CD and WF subscales that Herman and Polivy (1975) identified during the original development of the RS, is the most widely validated and frequently used conceptualization of the measure. This model has been supported by a variety of studies, including primarily normal-weight participants (Allison et al., 1992; Blanchard & Frost, 1983; Cole & Edelmann, 1987; Drewnowski et al., 1982; Heatherton et al., 1988; Lowe, 1984; Polivy et al., 1988; Ruderman, 1983). In most cases, Items 1, 5, 6, 7, 8, and 9 load on the CD factor, and Items 2, 3, 4, and 10 load on the WF factor (Blanchard & Frost, 1983; Drewnowski et al., 1982; Ruderman, 1983). Two factors often account for 50% to 60% of the variance. Herman and Polivy (1975) originally found the correlation among the two factors to be .48, while the more recent studies found subscale correlations ranging from .17 to .62.

Further evidence for the usual two-factor model was found by Allison et al. (1992), who performed orthogonal and oblique confirmatory factor analyses. The CD and WF factors accounted for 39% and 15% of the total variance, respectively. The original CD and WF scales correlated at .50.

Boerner et al. (2004) used structure equation modeling to conduct a confirmatory factor analysis on the RS. To facilitate the analysis, they combined items into parcels for factors with four or more items. The sample included 215 female and 214 male college students. The results indicated that the standard two-factor structure was a less than optimal fit using the comparative fit index (CFI = .85) but a fair fit using the root mean square error of approximation (RMSEA = .08). Similarly, in a series of factor analyses by Klem, Klesges, and Shadish (1990) on a sample of 229 college students (117 men, 112 women), the traditional two-factor model was only a fair fit with the data.

Occasionally, studies find more than two factors in the RS. Often, the results are attributed to the poor performance of certain specific items. van Strien et al. (2002) point out that there is generally poor consensus on the factorial assignment of Items 6 (splurging), 7 (thoughts about food), and 10 (history of overweight). The authors used maximum likelihood factor analysis to examine the RS responses from a sample of 209 female college students. The initial results suggested that a three-factor model fit the data the best, c^2 (35) = 13.65, p = .75. After oblique rotation, most items had high loadings on the first factor (36% of the variance), but the items from the WF subscale (2, 3, 4, 10) had the highest loadings on this factor. The five items from the CD subscale (1, 5, 7, 8, 9) and one item from the WF subscale (10) loaded highly on the second factor (9% of the variance). All items loaded negatively on the third factor (only 3% of the variance). Item 6 loaded highly on the first and third factors. When a two-factor solution was examined, Items 1, 6, and 10 were observed to load highly on both factors. The authors repeated their analysis after eliminating Items 6 and 10 due to their failure to load reliably on a single factor. Item 1 was kept because it was "considered central to the concept of dietary concern." The best-fit model included two factors, c^2 (13) = 12.55, with Items 2, 3, and 4 loading on the first factor (WF; 33% of the variance) and Items 1, 5, 7, 8, and 9 loading on the second factor (CD; 14% of the variance).

In a sample of 110 college students, Williams, Spencer, and Edelmann (1987) used principal components analysis to identify three factors with an eigenvalue greater than 1.4. The first factor included items primarily from the WF (1, 2 3, 4, 10) subscale, the second included items primarily from the CD subscale (5, 6, 8), and the third factor, labeled *attention to food intake,* included Items 2, 7, and 9. In this case, Item 2 loaded on the second and third factors. The three factors accounted for 27.7%, 21.4%, and 13.8% of the variance, respectively.

The findings of van Strien et al. (2002) and Williams et al. (1987) serve as a reminder that Herman and Polivy's conceptualization of restraint, as measured by the RS, includes several aspects of eating and attitudes, behaviors, and personal history that are related, but not perfectly so. Researchers who intended to measure the construct of restrain as conceptualized by Herman and Polivy need to recognize that the heterogeneity of constructs being assessed may be problematic. Those who desire a more "pure" (i.e., unidimensional) measure restraint are encouraged to use the restraint subscale from the Three Factor Eating Questionnaire or the Dutch Eating Behavior Questionnaire.

The developers of the RS intended it to be used as a single-factor measure (Polivy et al., 1988), and in most situations involving normal-weight samples, it should be used that way in primary analyses. Use of the total RS score will allow comparison with the majority of studies that have been conducted using the RS. However, the accumulated psychometric evidence suggests that the RS is multifactorial. van Strien et al. (2002) state that "use of total RS scores should be strongly discouraged" because the CD and WF subscales appear to measure qualitatively different constructs that may relate to outcomes such as disinhibited eating in different directions. Furthermore, the CD and WF subscales may interact in unpredictable ways. As such, it may often prove instructive to conduct secondary analyses that reanalyze data using the separate CD and WF subscales. If results replicate with one factor but not the other, it may provide valuable information about the source of the findings with the full scale.

Factorial Composition in Samples With a Significant Proportion of Overweight Participants

The two-factor model of the RS does not appear to be as reliable in samples composed primarily of overweight or eating-disordered participants. Most often, these studies report three or more factors (W. G. Johnson, Corrigan, Crusco, & Schlundt, 1986; W. G. Johnson, Lake, & Mahan, 1983; Lowe, 1984; Ruderman, 1983). Oblique factor rotation on samples including large numbers of obese participants often finds that Items 6 and 7 load on a third factor, possibly related to overeating. For example, Ruderman (1983) identified a four-four factor solution in a sample of 58 obese college students with a principal components factor analysis with orthogonal rotation. The factors consisted of a Weight Fluctuation dimension (25% of the variance), a Binge dimension (17% of the variance), a Tendency to Diet dimension (15% of the variance), and an Overconcern With Dieting dimension (12% of the variance). In addition, Lowe's (1984) exploratory principal components analysis found three factors with eigenvalues > 1.0. After oblique rotation, Items 1, 5, 8, 9, and 10 loaded on the first factor (29.3% of the variance), dubbed *dietary concern and weight history.* The second factor (28.3% of the variance), *weight fluctuation,* consisted of Items 2 to 4. Items 6 and 7 loaded on a third factor (17.6% of the variance).

The greater the proportion of overweight people in a sample, the more factors emerge (Ruderman, 1986). This factor instability may be a sign of differential validity or the result of restricted variance due to homogeneity of the sample. When a sample is homogeneous, the correlation coefficients among items are reduced, leading to an increased likelihood of the identification of additional factors in a factor-analytic study.

Factor Stability

A few studies have been conducted to test the factor stability of the RS. Blanchard and Frost (1983) found the factor structure of the RS to be stable across two samples of female college students. Tucker's (1951) congruence coefficient (CC) was above .99 for both factors, indicating

excellent factor stability. Allison et al. (1992) found that the CC for the RS factors for males and females was over .95. For random splits of the sample, the CC was over .99. A comparison of obese and nonobese subjects produced a CC of .96 for the CD factor and .92 for the WF factor. Boerner et al. (2004) used the guidelines described by Hoyle and Smith (1994) to test the factor stability of the RS for a sample of college men (n = 214) and women (n = 215). The authors conclude that the RS is invariant across gender.

Construct Validity: Convergent and Discriminant Validity

As opposed to other restraint scales that appear to measure actual dieting behaviors associated with caloric restriction (e.g., TFEQ, DEBQ), the RS appears to measure failed attempts at dieting (Heatherton et al., 1988). Researchers frequently consider the construct of restraint, *as measured by the RS,* to encompass both efforts at restricting food intake *and* episodes of overeating (van Strien, 1997). This conceptualization of restraint, as measured by the RS, was supported in analyses by van Strien et al. (2007), who used confirmatory factor analysis to examine the RS in relation to other measures of dieting, overeating, and body dissatisfaction in a sample of normal-weight (n = 349) and overweight (n = 409) females. A three-factor model was posited. The first factor, labeled *overeating,* consisted of the TFEQ disinhibition scale, the DEBQ emotional eating scale, DEBQ external eating scale, the Eating Disorder Inventory Revised (EDI-II) bulimic eating scale, and the question, "Have you ever had an eating binge, i.e., you ate an amount of food others would consider unusually large?" The second factor, labeled *dieting,* consisted of the DEBQ restraint scale, the TFEQ restraint scale, and the question, "Are you currently dieting?" The third factor, labeled *body dissatisfaction,* consisted of the EDI-II drive for thinness and body dissatisfaction scales. The confirmatory factor analyses were conducted at the level of scale scores rather than individual items. A model in which the RS loaded on all three factors was a better fit of the data than a model in which the RS loaded only on the dieting factor. The association of the RS with the overeating factor supports the conceptualization of the RS as a measure of unsuccessful dieting.

Further support for the RS as a measure of unsuccessful dieting comes from a study by K. K. J. Ferguson, Brink, Wood, and Koop (1992), who studied the individual RS item responses of a group of overweight participants in a dieting program. A group of 41 female and 41 male successful dieters was identified, who lost at least 5% of their body weight and maintained the loss for a year with no more than 5 lbs. regain. Unsuccessful dieters, including 32 women and 28 men, failed to meet these benchmarks. Unsuccessful dieters were more likely than successful dieters to endorse items related to overeating and food obsession, such as, "Do you eat sensibly in front of others and splurge alone?" and "Do you give too much time and thought to food?" On the other hand, unsuccessful dieters were less likely to endorse items related to restriction of food intake, such as, "How conscious are you of what you are eating?" This study is partly consistent and partly inconsistent with what Herman and Polivy's restraint theory would predict: Unsuccessful dieters were higher on disinhibiton items but *lower* on restriction items. According to Herman and Polivy, unsuccessful dieters should be higher on both because the continuing attempts to restrict presumably should be fueling the overeating.

Weight and Obesity Status

Given that the RS is associated with both efforts at caloric restriction *and* a propensity toward overeating, it is not surprising that researchers have found a variety of relationships

with weight and obesity status. Researchers have studied the relationship between the RS and weight primarily by correlating RS scores with body weight and body mass index (BMI), comparing the weight and BMI of restrained and unrestrained eaters, and comparing RS scores among normal-weight and overweight participants. Drewnowski et al. (1982) found a relationship between only the WF subscale and percentage overweight. Drewnowski et al. also found that overweight participants scored higher than normal-weight participants on the total RS and the WF subscale but not the CD subscale. Because greater weight fluctuations in overweight individuals could stem from biological characteristics of adipose tissue per se (rather than from repeated periods of weight loss dieting and disinhibition-induced weight regain), Drewnowski et al. suggested that the RS may not be an appropriate measure of restrained eating in overweight individuals. However, Lowe (1984) found that CD (r = .41) but not WF (r = –.01) was related to overweight status in a sample of 217 college students (96 men, 118 women, 3 unknown). The discrepancy between the Lowe and Drewnowski et al. findings is likely the result of a greater proportion of overweight participants in the Drewnoswki sample. This interpretation is supported by Allison et al. (1992), who found that obese participants (n = 78) obtained significantly higher scores on the RS and the WF subscale but not CD.

In two studies, Ruderman (1983, 1985) found correlations of .37 and .38 between RS scores and percentage overweight. In a study comparing overweight and nonoverweight participants, Klem, Klesges, Bene, and Mellon (1990) found that overweight participants obtained significantly higher scores on the CD and WF subscales, as well as on the total RS. In a sample of 358 adults (201 men and 157 women), de Castro (1995) found that higher RS scores were associated with higher body weights. Similarly, a Portuguese translation of the RS was significantly correlated with BMI in a sample of patients suffering from anorexia nervosa or bulimia nervosa (r = .38) and non-eating-disordered controls (r = .43; Scagliusi et al., 2005). Lowe (1984) found that restrained eaters had greater relative weights than unrestrained eaters, even though all participants were within the normal weight range.

The RS failed to prospectively predict changes in body weight in three studies involving college students (Klesges, Klem, Epkins, & Klesges, 1991; Lowe et al., 2006; Tiggemann, 1994). However, Klesges et al. (1992) found that RS scores predicted weight gain among adult women but not men over a 1-year period when the relationship was analyzed in a multiple linear regression, including other physiological, demographic, and activity variables. Williamson et al. (2007) reported that RS scores increased during a weight loss intervention, but changes in RS were not correlated with relative energy balance during the diet.

There appears to be a relationship between the RS and body weight. However, the relationship is not consistent across samples and may be artificially inflated among overweight and obese individuals. Given that nearly all literature on the RS has involved primarily normal-weight individuals, that overweight restrained eaters and dieters do not behave like those of normal weight (Lowe et al., 1991; Ruderman, 1986), and that the RS has weaker psychometric properties in overweight individuals, the RS is not well suited as a measure of restrained eating in overweight samples.

Naturalistic Food Consumption

Several authors have attempted to find a relationship between the RS and measures of naturalistic food consumption. However, most of these studies rely on self-reported dietary intake via food diaries, which have poor validity in general (Bandini, Schoeller, Dyr, & Dietz, 1990; Lichtman et al., 1992; Livingstone, Prentice, & Strain, 1990; Prentice et al., 1986), but especially

among overweight samples (Lichtman et al., 1992; Prentice et al., 1986) and restrained eaters (for a review, see Maurer et al., 2006). Both of these groups tend to underreport food intake to a significantly greater degree than unrestrained normal-weight individuals.

Laessle, Tuschl, Kotthaus, and Pirke (1989) failed to find a correlation between RS ($r = -.04$) and mean caloric intake over a 7-day period in a sample of 60 normal-weight women. Similarly, de Castro (1995) found no relationship between total caloric intake and RS over a 7-day period in a sample of 201 male and 157 female adult participants. In a study by French et al. (1994), RS score was not related to caloric intake over a 6-month period, as measured by the Block Food Frequency Questionnaire (FFQ; Block et al., 1986). All three of these studies relied on self-reported intake. The fact that restrained eaters are more likely to underreport their actual food intake could be masking a tendency toward greater intake in restrained eaters. Consistent with this speculation are findings indicating that measures of restrained eating prospectively predict weight gain rather than weight loss (Stice et al., 2004).

Eating Disorders and Psychopathology

The creators of the RS have suggested that dietary restraint and eating-disordered attitudes and behaviors are inherently related and have gone so far as to say that the type of dieting that is measured by the RS can lead to the development of eating disorders (Heatherton & Polivy, 1992; Polivy & Herman, 1985). A variety of cross-sectional studies support this claim. Ruderman and Grace (1987) found that the RS was correlated with the BULIT (Smith & Thelen, 1984), a measure of bulimia, in a sample of 108 women. The partial correlation between the BULIT and the CD subscale of the RS was still statistically significant when WF subscale scores were controlled. However, the relationship between WF and the BULIT was nonsignificant when the CD scores were controlled. In a sample of college students (Boerner et al., 2004), the RS total score was significantly correlated with the BULIT-R (Thelen, Farmer, Wonderlich, & Smith, 1991) among both men ($r = .56$, $n = 214$) and women ($r = .69$, $n = 215$). In addition, scores for both men ($r = .46$) and women ($r = .64$) were correlated with a measure of anorexic symptomatology, the Eating Attitudes Test (EAT; Garner & Garfinkel, 1979). Using a Portuguese translation of the RS, Scagliusi et al. (2005) found that bulimics ($n = 24$) scored significantly higher on the RS than anorexics ($n = 15$), who obtained significantly greater scores than non-eating-disordered college students ($n = 57$). Prussin and Harvey (1991) compared a subsample of 38 individuals meeting *DSM-III-R* criteria for bulimia to 136 non-eating-disordered participants in a sample of normal-weight female runners. Bulimic participants had significantly higher RS scores. Bourne, Bryant, Griffiths, Touyz, and Beaumont (1998) found that the RS and its subscales were significantly correlated with greater frequency and intensity of disordered eating behaviors, as measured with the Eating Behavior Rating Scale (Wilson, Touyz, Dunn, & Beumont, 1989), during a video-recorded test meal. Griffiths et al. (2000) found significant relationships between the RS and abnormal eating attitudes and general dissatisfaction with one's life in a sample of 82 college students.

Prospective studies have confirmed that elevated RS scores predict the future onset of binge eating (Stice, Killen, Hayward, & Taylor, 1998) and bulimic pathology (Killen, Taylor, Hayward, & Wilson, 1994; Killen et al., 1996). In a sample of 967 adolescent girls who were followed over a 4-year period, Killen, Hayward, Wilson, and Taylor (1994) found that girls who developed bulimic symptoms had greater scores on both the CD and WF subscales of the RS at baseline compared to girls who remained asymptomatic. In a similar study of 543 female high

school students, Stice et al. (1998) reported that RS scores at baseline predicted onset of objective binge eating, subjective binge eating, and purging. Two items referring to binge eating were removed from the RS for this analysis because of concerns regarding criterion confounding, which are discussed below.

Scores on the RS are clearly associated with measures of eating-disordered attitudes and behaviors. This is not surprising since dieting is a cardinal feature of both anorexia and bulimia nervosa, and overeating is a cardinal feature of bulimia nervosa. In addition, there is some evidence that RS scores are associated with depression and general dissatisfaction with life. However, Stice et al. (1997) suggest that the relationships observed between the RS and measures of eating-disordered symptomatology are the result of criterion confounding of the RS, which includes items related to disinhibited eating, a close relation of binge eating. When these items were removed (Items 6 and 8), the relationship between the RS and measures of disordered eating were significantly reduced among a sample of 117 female college students. The relationships were further weakened when items pertaining to weight fluctuation (which may create an artificial relationship between the RS and measures of eating-disordered symptomatology) were removed. The authors' argument for criterion confounding of the RS is strengthened by the fact that the DEBQ-R, which does not included items related to weight fluctuation or disinhibited eating, did not show equivalent relationships with measures of disordered eating.

Because the RS and other measures of restrained eating have been linked to the development of unhealthy eating behaviors, it is now widely accepted that "dieting" plays a causal role in the onset of eating disorders (e.g., Hawkins & Clement, 1984; Heatherton & Polivy, 1992; Polivy & Herman, 1985). In rare cases involving radical dieting and extensive weight loss to subnormal levels, there is reason to believe that such a connection exists (e.g., Butryn & Wadden, 2005). However, experimental evidence suggests that prescribed diets involving gradual weight loss reduce binge eating in normal-weight and overweight individuals (for a review, see Stice et al., 2004). This evidence, combined with studies indicating that restraint scales do not reflect hypocaloric dieting (Stice et al., 2007), seriously questions the prevalent assumption that garden-variety dieting helps cause eating disorders.

Susceptibility to Response Sets

Historically, restrained eaters were thought to be motivated by a desire to attain a thin body to conform to socially defined standards for attractiveness (Polivy & Herman, 1987). Furthermore, some items on the RS, especially those related to overeating, may be embarrassing to endorse. Thus, it seems plausible that the RS may be influenced by social desirability bias, which is the inclination to present oneself in a manner that will be viewed favorably by others. Several researchers have tested this theory by correlating the RS with measures of social desirability responding. Most measures of social desirability responding present participants with a list of behaviors that are either socially desirable but infrequently practiced or frequently practiced but socially undesirable. Attempts to "fake good" are indicated by endorsement of the former type of behavior and denial of the latter type. The Minnesota Multiphasic Personality Inventory (MMPI) L, or "lie" scale, is possibly the most well-known measure of social desirability responding. The items comprising the Edwards Social Desirability Scale (Edwards, 1957), and some items from the Marlowe-Crowne Social Desirability Scale (MCSD; Crowne & Marlowe, 1964) were taken from the MMPI.

W. G. Johnson et al. (1983, 1986) found small and nonsignificant correlations between the RS, the MMPI Lie scale, and the MCSD for bulimics, obese nondieters, and "normals." However, the relationship between the RS and the MMPI Lie scale (r = –.33), as well as the RS and MCSD (r = –.51), was moderate and negative for a sample of 27 obese dieters (W. G. Johnson et al., 1983). Ruderman (1983) found the opposite; the relationship between the RS and the Eysenck Lie Scale was stronger for nonobese participants (r = –.70) than obese participants (r = –.13). Other studies have found small and nonsignificant correlations between the RS and the Edwards Social Desirability Scale (Kickham & Gayton, 1977) and the RS and MCSD among normal-weight participants (Corrigan & Ekstrand, 1988; Ruderman, 1983) and obese participants (Ruderman, 1983). In a subset of participants (n = 73), Allison et al. (1992) found that the RS correlated with the MCSD (r = –.27) and the Edwards Social Desirability Scale (r = –.05). The authors also found that RS items that were rated as more desirable were endorsed more frequently. In the same study, when participants were instructed to "create the most favorable impression you can," scores on the RS were low (mean = 8.75). When instructed to "create the worst possible impression," the mean score was very high (mean = 30.65).

Generally, the relationship between the RS and social desirability scales is negative, meaning that high scores on the RS are associated with relatively elevated endorsement of socially undesirable behavior. These findings present an interesting contrast to restraint theory, which suggests that restrained eaters' behavior is motivated by a desire to attain a more socially desirable appearance. Regardless, the RS is transparent and can easily be "faked" good or bad. Finally, McCrae and Costa (1983a, 1983b) point out that correlations between a psychometric instrument and measures of social desirability responding should not necessarily be taken as a sign of invalidity of the instrument. It is generally undesirable to have a measure correlate with socially desirable motives, unless such a relationship can be argued to be part of the construct the measure is supposed to assess. In the case of the RS, the creators of the scale explicitly state that individuals who score highly on the measure are presumed to be highly influenced by socially dictated standards for appearance (Polivy & Herman, 1987).

Predictions of Laboratory Behavior

The RS is well known for its ability to predict disinhibited eating in laboratory studies using the preload paradigm (Herman & Polivy, 1984; Lowe, 1993). In these studies, participants are typically designated as restrained or unrestrained eaters based on the median score of the RS. Half of each group will be assigned to consume a high-calorie preload, such as a milkshake, before they participate in a "taste test" of palatable food, such as ice cream. The outcome measure is the amount of food consumed during the taste test, which is surreptitiously monitored by the experimenter. Unrestrained eaters typically compensate for a preload by consuming fewer calories in the preload than in the no-preload condition. Restrained eaters show the opposite trend: They will show evidence of disinhibited eating and consume somewhat *more* after than in the absence of a preload. This pattern of findings is typically observed only when dietary restraint is measured with the RS but not other measures such as the TFEQ or DEBQ (Lowe, 1993).

Notably, a caloric preload is not the only stimulus that will lead to disinhibited eating. Emotional distress (Herman & Polivy, 1980), threat of electric shock (Herman & Polivy, 1975), and increased cognitive load (Ward & Mann, 2000) also result in disinhibition. Furthermore, restrained eaters will exhibit disinhibited eating when they are led to believe that they have

consumed a high-calorie preload when in fact the preload they consumed was low in calories (e.g., Heatherton, Polivy, & Herman, 1989). Thus, disinhibited eating seems to occur when restrained eaters believe that their efforts at caloric restriction have been "blown" or when they are distracted from their efforts at restraint by an engrossing or distressing stimulus.

The trend toward disregulation of food intake by restrained eaters was also observed in a study by Westerterp-Plantenga, Wouters, and ten Hoor (1991) in which 6 obese and 18 normal-weight women were served a four-course meal. Participants were allowed to eat as much as they wished during the second course, but the amount of food served during the other three courses was fixed. Eating behavior was observed, and the amount of food eaten was surreptitiously measured by a scale under the participant's plate. Participants who were low on the RS scale showed a decreased rate of intake following the first course. Restrained women showed a pattern of progressive linear intake across the meal. This result may reflect the same process (lack of response to eating what is normally a satiating amount of food) as observed in preload studies, even though the indicator was different (rate of eating over the meal).

The relationship between the RS and eating behavior observed in the laboratory is complex (Lowe, 1993). A sizable minority of studies have failed to find evidence of disinhibited eating in restrained eaters (e.g., Ouwens, van Strien, & van der Staak, 2003; van Strien, Cleven, & Schippers, 2000), while some have found that the effect of disinhibition is better accounted for by other constructs such as attributional style (Rotenberg & Flood, 2000). In addition, van Strien et al. (2002) found that the WF and CD subscales interacted with the preload in opposite directions in the prediction of food intake during the taste test, suggesting that the component parts of Herman and Polivy's Restraint Scale may be differentially related to behavioral outcomes. Finally, as with other aspects of restrained eating, the outcome of laboratory studies seems partly dependent on the participants' weights. van Strien et al. (2007) note that the disinhibition effect has never been observed in overweight restrained eaters. This observation reinforces the recommendation that the RS not be used in overweight samples.

While some of the laboratory studies cited here seem to suggest that restrained eaters eat less than unrestrained eaters in the absence of a disinhibiting stimulus (Herman & Polivy, 1984), a series of studies by Stice and colleagues (Stice et al., 2004, 2007) strongly suggest that such laboratory-based findings of reduced eating by restrained eaters in the laboratory do not generalize to their food intake outside the laboratory.

Readability

The reading level of the RS has been estimated to be between the fourth and ninth grades (Allison & Franklin, 1993).

Stunkard and Messick's TFEQ-R Scale

Description

The Three Factor Eating Questionnaire (TFEQ), also known as the Eating Inventory (Stunkard & Messick, 1988), was created by Stunkard and Messick (1985) in response to a developing awareness of the limitations of the RS. The authors expressed concerns with

regard to the content of the RS and its construct validity. In regards to the content of the RS, the authors point out that, while the RS was not designed to measure the behavior of overweight and obese persons, its creators had suggested that the RS measured the construct of dieting as *separate* from the construct of overweight. Furthermore, they indicated that the cause of many behaviors associated with obesity was a history of dieting per se (Hibscher & Herman, 1977). However, it became increasingly apparent that the RS was indeed influenced by obesity. Some studies reported that overweight restrained eaters did not show evidence of disinhibited eating as did normal-weight restrained eaters. Furthermore, the RS contains items related to weight fluctuation that may artificially inflate the scale scores of persons suffering from overweight and obesity. Finally, the relationships that researchers reported for the RS and various outcome measures such as food consumption varied in strength and even direction, and the relationships seemed to vary by obesity status. Herman and Polivy's hypothesis that restraint accounted for the eating behavior of obese individuals was not supported by reports that restrained obese individuals did not demonstrate counterregulatory eating (Ruderman, 1986).

In response to these concerns, as well as the desire for a measure that would be more reliably related to food intake in normal-weight and obese persons, Stunkard (1981) and later Stunkard and Messick (1985) developed the restraint scale of the TFEQ (TFEQ-R). The first version of the TFEQ borrowed several items from the RS and Pudel et al.'s (1975) Latent Obesity Questionnaire, and 17 original items were also included. The variety of questions included in the scale reflects Stunkard and Messick's intention to capture several facets of eating behavior, including but not limited to dietary restraint.

The original 67-item scale was administered to a sample of 220 participants, including both genders and persons of both obese and normal weight. An exploratory factor analysis including all participants suggested three factors, representing behavioral restraint, lability in behavior and weight, and hunger. The results were essentially equivalent when separate factor analyses were conducted for men and women, as well as three groups of participants who were ostensibly low, medium, and high on restraint.

On the basis of these preliminary results, the authors modified some items and added others in an effort to capture more accurately the constructs measured by each of the newly identified factors and to heighten the distinctiveness of each factor. A new sample, consisting of 53 (7 men and 46 women) participants in the same intensive weight loss program and 45 (5 men, 13 women, and 27 of indeterminate gender) completed a questionnaire comprising 93 items, including those that were unchanged, modified, and newly written. Of those, 58 items were selected for inclusion in the final version of the TFEQ. The items in the final measure were selected because of significant partial correlations with their provisional factors, while holding the other two subscales constant. Finally, the subscales were given new names: Cognitive Control of Eating (Factor I), Disinhibition (Factor II), and Susceptibility to Hunger (Factor III). Cronbach's alpha was .92, .91, and .85 for Factors I, III, and III, respectively. A correlation of –.43 was found for Factors I and II, –.03 for Factors I and III, and .42 for Factors II and III. Although the scale was originally published as the Three-Factor Eating Questionnaire (Stunkard & Messick, 1985), it is now published by the Psychological Corporation as the *Eating Inventory* (Stunkard & Messick, 1988). For the purposes of the present chapter, we shall confine our discussion mainly to the Restraint Factor scale and shall refer to the restraint scale of the TFEQ as the TFEQ-R.

Sample

As described in the previous section, a preliminary set of items was tested on a sample of 97 men and 123 women. The sample consisted of 78 "dieters" who were members of an intensive weight loss group, 62 nonobese "free eaters" who were selected by the dieters, and 80 persons who were chosen by the dieters for geographic proximity. The ages of the participants ranged from 17 to 77 years with a mean of 44.

A second sample of 53 dieters (7 men and 46 women) and 45 free eaters (5 men, 13 women, and 27 of indeterminate gender) was used to refine the instrument. As before, the free eaters were nominated by the dieters, who were recruited from an intensive weight loss program. This second sample was used to identify the norms in the next section.

Norms

Means, sample sizes, and standard deviations for participant groups on the TFEQ-R are presented in Table 5.4. As with other measures of restraint, studies often report lower TFEQ-R scores for men than women (e.g., Bellisle et al., 2004; de Castro, 1995). Stunkard and Messick (1988) suggest tentative TFEQ-R guidelines of 0 to 10 as "low average," 11 to 13 as "high," and 14 or more as "clinical range." Care should be taken when attempting to classify persons into high- or low-restraint groups, as TFEQ-R scores differ by gender and nationality. Furthermore, scores should be interpreted in the context of the other characteristics of the responder. For example, a low TFEQ-R score in an obese person with obesity-related health problems may be a cause for concern, whereas a high restraint score in a thin woman could be problematic.

It should also be noted that researchers sometimes change the dichotomized response format of the true/false items in the TFEQ-R to a 4-point response scale. This practice seems especially common in twin studies of the genetic component of eating behaviors (e.g., Neale, Mazzeo, & Bulik, 2003; Tholin, Rasmussen, Tynelius, & Karlson, 2005). While this practice may facilitate studies of heredity, the TFEQ-R scores reported in these studies are not directly comparable to studies using the standard scoring rubric.

Reliability

Internal Consistency

As can been seen in Table 5.2, Cronbach's alpha for the TFEQ-R is routinely reported to be at or greater than .80. Unlike the RS, the TFEQ appears to be equally reliable for normal-weight and obese persons.

Test-Retest

Stunkard and Messick (1985) cited an unpublished manuscript by Ganley that reported a test-retest correlation over a 1-month interval to be .93. Allison et al. (1992) found test-retest

Table 5.4 Mean Three-Factor Eating Questionnaire–Restraint Scale (TFEQ-R) Scores Reported in the Literature

Participants	*Author*	*n*	*Mean*	*SD*
Unrestrained eaters	Stunkard and Messick (1985)	62	6.0	5.5
Swedish control group	Bjorvell et al. (1986)[a]	58	9.8	4.2
Chilean university students	Lolas (1987)[a]	88	7.7	5.1
U.S. control sample	Ganley (1986)[a]	30	11.0	5.3
American adult men	French, Jeffery, and Wing (1994)	99	5.9	4.2
American adult women	French et al. (1994)	103	9.1	4.2
Postmenopausal American women At baseline At 4-year follow-up	Hays, Bathalon, Robenoff, McCrory, and Roberts (2006)	 36 36	 10.6 9.0	 6.9 5.5
American adults	Williamson et al. (2007)	46	7.8	4.1
Japanese high school girls	Nogami (1986)[a]	243	5.6	3.7
Female White American college students	Atlas, Smith, Hohlstein, McCarthy, and Kroll (2002)	300	10.4	5.4
Female African American college students		200	8.9	5.3
American college students Females only Males only	Allison et al. (1992)	901 617 282	9.0 10.26 6.1	5.8 5.6 5.1
American male college students	Boerner, Spillane, Anderson, and Smith (2004)	214	4.7	4.7
American female college students	Boerner et al. (2004)	215	8.2	5.7
Australian female college students	Ricciardelli, Tate, and Williams (1997)	172	15.9	8.4
Japanese nursing students	Nogami (1986)[a]	270	6.3	3.6
German women	Laessle, Tuschl, Kotthaus, and Pirke (1989)	62	6.5	4.7
German women in a weight reduction program	Westenhoefer (1991)	46,132	13.1	4.3
German men in a weight reduction program	Westenhoefer (1991)	8,393	10.6	4.7

a. Reported in Stunkard and Messick (1988).

reliability to be .91 over a 2-week span. Bond, McDowell, and Wilkinson (2001) reported a test-retest coefficient of .81 over 1 year.

Validity

Content Validity

Factorial Composition. While the focus of this chapter is measures of restraint, findings involving the other two TFEQ subscales are reviewed below because they can help shed light on the domain assessed by the TFEQ-R. Factor analyses of the full TFEQ, including items from all three subscales, typically find that a three-factor solution fits the data well. Stunkard and Messick (1985) conducted several factor analyses during development of the measure, with the express intention of creating distinct subscales. Little variation in the factor structure was found between dieters in a weight loss program who were ostensibly restrained eaters and neighbors of the dieters who were ostensibly moderately restrained. However, the factor structure for a group of "free eaters" was slightly less simple, possibly because of infrequent endorsement of items related to restraint and disinhibition. Regardless, the restraint factor (Factor I) was robust across all groups. Highly similar results were obtained by Hyland, Irvine, Thacker, and Dan (1989) and Ganley (1988).

Boerner et al. (2004) used structural equation modeling to conduct a confirmatory factor analysis of several measures of eating attitudes and behaviors simultaneously. Items from the subscales of each measure were combined into item parcels to facilitate analysis. The authors found that the typical three-factor model fit the TFEQ very well. Similar results were obtained by Atlas, Smith, Hohlstein, McCarthy, and Kroll (2002). In contrast, Mazzeo, Aggen, Anderson, Tozzi, and Bulik (2003) tested three models of the TFEQ using two types of confirmatory factor analysis and found that none of the models produced an acceptable fit of the data. However, the authors used a modified TFEQ that excluded 15 items and altered the response option for some other items. It is unclear to what degree the results reported in this study may have been affected by Mazzeo et al.'s manipulation of the TFEQ.

Of greater relevance to the study of restrained eating are studies that focus more specifically on the 21 items of the TFEQ-R. Ricciardelli and Williams (1997) examined the factor structure of the TFEQ-R. The sample consisted of 144 female college students. A principal components analysis with varimax rotation identified three factors. The first factor, accounting for 33.5% of the variance, included six items and was labeled *emotional/cognitive concerns for dieting.* The second factor contained seven items, accounted for 7.8% of the variance, and was labeled *calorie knowledge.* The third factor was made up of five items, accounting for 6.6% of the variance, and was labeled *behavioral dieting control.* Three items failed to load on any of the factors. Ricciardelli and Williams suggested that Factors I and III are similar to the constructs of cognitive restraint and behavioral restraint that have been identified in the literature on problem drinking. They conclude that Factor III may be a better measure of successful dieting than the total TFEQ-R, as Factor III was negatively correlated with BMI.

Westenhoefer (1991) identified two highly correlated sources of variance in the TFEQ-R using a variant of discriminant analysis. In a sample of 46,132 female and 8,393 male Germans in a weight loss program, factors were identified representing "Flexible" control and "Rigid"

control over eating. Persons scoring highly on the rigid control scale were characterized by a dichotomized, "all-or-nothing" approach to eating. They reported dieting frequently but did not seem to follow any specific plan. On the other hand, individuals scoring highly on flexible control reported eating more slowly, taking smaller helpings, and controlling their eating by using situation-specific guidelines rather than inflexible rules. Rigid control was associated with high disinhibition, whereas flexible control was linked to low disinhibition.

Allison et al. (1992) conducted a principal components factor analysis on the TFEQ-R responses of 901 college students. While the Minimum Average Partial (MAP) test (Zwick & Velicer, 1986) suggested a one-factor solution, and goodness-of-fit indices were good to fair for this model, a two-factor solution similar to that found by Westenhoefer (1991) was eventually retained. Catell's scree test and the Guttman-Kaiser eigenvalues > 1 rule each suggested a two-factor solution. Varimax rotation was attempted but later abandoned when an oblique rotation yielded a simpler factor pattern. The two factors seemed to represent a cognitive dimension (35% of the variance) and a behavioral dimension (6% of the variance) of restraint. However, the correlation between the factors was high (r = .56). The authors concluded that the TFEQ-R contains two highly correlated primary factors that can be considered nested within a broader secondary factor.

Gorman et al. (1993) conducted a further analysis of the findings reported in Allison et al. (1992). The high correlation between the two factors and the substantially greater endorsement of items in the cognitive restraint factor as compared to the behavior restraint factor led the authors to consider alternative methods of analyzing the data. Psychometric research has shown that conventional linear factor analysis techniques will often produce spurious factors when items differ considerably in their endorsement rates (G. A. Ferguson, 1941; Gibson, 1967; Horst, 1965; McDonald & Ahlawat, 1974). Thus, Gorman et al. reanalyzed the data using nonlinear techniques, including multidimensional scaling and Rasch model scaling (Hambleton, Swaminathan, & Rogers, 1991), that alleviate the biases of traditional methods. The results suggested that the TFEQ-R items form a continuum that begins with relatively common thoughts of reducing eating and ends with overt, deliberate, but relatively rare actions to reduce eating.

Taken together, these findings suggest that the TFEQ-R performs well as a unidimensional measure of restrained eating but that it can also be further bifurcated into a cognitive and a behavioral component. There is evidence that these two components may form a continuum ranging from typical thoughts of reducing intake to actual behaviors at limiting consumption that are rarely followed through with. While further research involving nonlinear analytic techniques is needed to strengthen this conceptualization of the TFEQ-R, it appears that these results dovetail nicely with recent data suggesting that restrained eaters on the TFEQ-R do not actually reduce their food intake below their energy needs, even though they may wish they could do so (Stice et al., 2004, 2007).

Factor Stability

Allison et al. (1992) found that Tucker's CC was high for random splits of the subject sample (CC > .97) but only modest (CC < .90) when comparing obese and normal-weight participants and low when comparing across gender (CC < .90). However, Boerner et al. (2004) found that the TFEQ-R was invariant across gender using the steps described by Hoyle and Smith (1994) for testing measurement invariance. Atlas et al. (2002) found that the TFEQ-R performed equally well for African Americans as Whites. More research is needed to compare the performance of the TFEQ-R in obese and normal-weight participants.

Construct Validity: Convergent and Divergent Validity

Relationships Among the TFEQ Subscales

Stunkard and Messick intended for the subscales of the TFEQ to be conceptually and empirically distinct. For the most part, this goal seems to have been accomplished, although there is notable overlap between the TFEQ-R and the other subscales, in some reports. For example, Atlas et al. (2002) found a moderately strong correlation between the TFEQ-R and the TFEQ Hunger subscale for both White (r = .74) and African American (r = .77) women. However, the correlation between the TFEQ-R and the TFEQ Disinhibition subscale was substantially stronger for White (r = .47) than African American (r = .05) women. Despite the correlations that have been observed in such studies, the TFEQ subscales were never intended to be combined into a single "total" score, and there is no evidence that such an amalgam has any theoretical or empirical utility.

The TFEQ subscales appear to relate to each other differently for obese and normal-weight individuals but similarly across gender. Bellisle et al. (2004) studied these relationships in a sample of 2,509 adults of both genders and varying weights. While correlations between the TFEQ-R and the Disinhibition and Hunger subscales were positive in the lowest BMI groups (i.e., BMI < 27), the relationship became increasingly more negative as BMI increased. In persons with BMI greater than 45, TFEQ-R was moderately negatively correlated with the Disinhibition subscale in women and men. In the same BMI category, the relationship between TFEQ-R and Hunger was r = –.30 for women and r = –.12 for men. Similar results were found by Foster et al. (1998), who reported moderately negative correlations between TFEQ-R and Disinhibition, as well as between TFEQ-R and Hunger, among overweight women seeking behavioral treatment for weight loss. In a sample of U.S. college students, Boerner et al. (2004) found that the TFEQ-R and Disinhibition subscale was moderately positively correlated among men and women. The correlation between TFEQ-R and Hunger was very weak for men and women. Similarly, in two samples of Dutch female college students, TFEQ-R was correlated with Disinhibition at r = .36 and r = .42 (Ouwens et al., 2003; van Strien et al., 2000, respectively). In a study by van Strien et al. (2007), a significant difference was observed in the correlation between TFEQ-R and Disinhibition for normal-weight (r = .41) and overweight (r = .07) subsamples. For people in the normal weight range, it may be that people with low Disinhibition or Hunger scores have very low risk for weight gain (they may be "naturally thin"), whereas those with higher scores may attempt to counter their chronic vulnerability to overeating and weight gain by being more restrained. Among those already obese, most may already be frequently overeating relative to their normal-weight peers, but those who are currently restraining their eating are (at least temporarily) reducing their vulnerability to this overeating.

Weight and Obesity Status

During the measure development process, Stunkard and Messick (1985) found a correlation of .20 between restraint and weight. Since then, a variety of relationships have been reported. Allison et al. (1992) found no significant differences between obese and nonobese participants on the TFEQ-R. Ricciardelli and Williams (1997) reported that the TFEQ-R correlated with BMI (r = .25), previous dieting (r = .64), and current dieting (r = .65) in a sample of female college

students. Beiseigel and Nickols-Richardson (2004) found that a subgroup of normal-weight college women with high scores on the TFEQ-R possessed more fat mass (as measured by dual-energy X-ray absorptiometry) and had a higher body fat percentage than a subgroup of women with low restraint scores.

When a French translation of the TFEQ-R was administered to 1,554 participants, 955 of whom were in the obese range, the TFEQ-R scale was positively associated with BMI in men but not in women (Bellisle et al., 2004). Obese and nononbese women did not differ significantly on the TFEQ-R. Furthermore, being obese as a child and/or adolescent was generally associated with more intense restraint, disinhibition, and hunger in adults, whether or not the subject was still obese at the time of the test. The authors conclude that some level of restraint may allow some children to grow out of obesity. This study was conducted with obese persons and their first-degree relatives, so the results may not be generalizable to persons with no family history of obesity.

De Lauzon-Guillain et al. (2006) studied the relationship between eating behavior and weight gain in a community sample of 466 adults and 271 adolescents over a 2-year period in France. At baseline, a French translation of the TFEQ-R was positively associated with BMI in normal-weight participants but not overweight adults. While TFEQ-R scores did not predict changes in adiposity, a higher initial BMI was associated with a larger increase in TFEQ-R. Similarly, Hays, Bathalon, Roubenoff, McCrory, and Roberts (2006) examined predictors of weight change in a sample of 36 nonobese postmenopausal women in a 4-year longitudinal study. Hunger was the only TFEQ subscale that predicted weight gain.

While the previous studies found either a positive relationship or no relationship between TFEQ-R and body size, Westenhoefer, Stunkard, and Pudel (1999) found that the TFEQ-R was negatively associated with BMI in both male and female Germans in a computer-assisted weight loss program. TFEQ-R was also positively associated with successful weight loss. Although not discussed by the authors, there were also apparently significant interactions between TFEQ-R and Disinhibition, such that the antiobesity effects of restraint were stronger at higher levels of disinhibition. This pattern of results was also observed by Williamson et al. (1995). These results are consistent with the previously mentioned argument that dietary restraint may be a desirable characteristic in already overweight individuals.

Westenhoefer et al. (1999) further parsed their results by the flexible and rigid control subscales developed by Westenhoefer (1991). These analyses revealed that rigid control is associated with increased Disinhibition and higher BMI, whereas flexible control is associated with lower Disinhibition and lower BMI. Furthermore, successful weight losers had more flexible control at the beginning of the program and increased their flexible control scores during the program, whereas less successful participants had lower scores at the beginning and did not increase them during the program. Differences for rigid control, while statistically significant, were considerably smaller. The authors conclude that flexible control, but not rigid control, is associated with successful weight reduction. However, just as the potential causal association between restraint and overeating is open to debate (e.g., overeating may increase restraint, not vice versa), so is the causal status of rigid and flexible dieting. It is possible that flexible dieters are able to be flexible because their overeating tendencies are not as severe, whereas rigid dieters have learned that they can only control their eating by employing more definitive dieting rules.

Generally, the TFEQ-R seems to be linked with successful weight loss. In addition to the studies described previously, Foster et al. (1998) found that weight loss treatment was associated with significant increases in restraint and decreases in disinhibition and hunger. Before treatment,

higher restraint scores were associated with lower body weights, and greater increases in restraint were correlated with greater weight losses. In a study of 46 adults (26 men and 20 women) seeking weight loss treatment, TFEQ-R scores increased significantly in treatment groups but not the control group (Williamson et al., 2007). Notably, of several measures of dietary restraint, the TFEQ-R was the only measure to be correlated with energy balance (as measured by a combination of doubly labeled water and change in body composition). However, it was the *change* in TFEQ-R, not its absolute value, that was associated with energy balance. Increases in TFEQ-R were associated with an energy deficit. Very little or no change in TFEQ-R was associated with energy excess. TFEQ-R is also related to weight maintenance. Westerterp-Plantenga, Kempen, and Saris (1998) found that participants who successfully maintained weight loss following a very low-calorie diet experienced greater increases in TFEQ-R during the diet, as compared to participants with poorer weight maintenance, who did not experience as great an increase in TFEQ-R while dieting.

TFEQ-R scales have also been linked to the construct of weight suppression (i.e., the difference between current and highest ever weight). de Castro (1995) reported an interaction between TFEQ-R and gender in the prediction of weight suppression in a sample of 201 male and 157 female adults. When participants were trichotomized based on their TFEQ-R scores, the current weights of high-restraint men were 10.5% below their highest weights, whereas the current weights of the moderate- and low-restraint groups were closer to their highest weights (5.2% and 6.6% below their highest weights, respectively). This pattern was not observed for women, who were 6.9% below their highest ever weight in all three restraint groups.

Naturalistic Food Consumption

de Castro (1995) reported that highly restrained eaters had significantly lower self-reported caloric intake than dieters with low restraint. The differences resulted from significantly lower intakes of fat and carbohydrate in restrained eaters (although the usual cautions about underreporting in restrained eaters apply). In addition, overall daily intakes were less variable with higher levels of restraint. Participants high in restraint had lower deprivation ratios but not satiety ratios. This suggests that highly restrained participants ate significantly less than unrestrained eaters relative to their period of premeal deprivation than did the less restrained subjects, but there was no differential effect of meal size on time to next meal.

In a study of food intake and physical activity, French et al. (1994) found that women who scored highly on the TFEQ-R reported significantly lower caloric intake, lower percent calories from sweets, and less frequent sweets consumption than women with low TFEQ-R scores. Men with high TFEQ-R scores reported a significantly greater percentage of calorie intake from protein and carbohydrate and less frequent consumption of beef, pork, whole milk, and sweets.

Lahteenmaki and Tuorila (1995) studied the relationship between the TFEQ-R and the desired use and liking of a variety of foods in a sample of 253 women and 11 men attending Weight Watchers in Finland. The TFEQ-R was negatively related to the reported use of some food groups such as fruit-based sweet foods, butter, margarine, and regular-fat cheese but not to their desired use or liking. Beiseigel and Nickols-Richardson (2004) found that college women who score highly on the TFEQ-R consumed more servings of fruits and vegetables per day compared to women in a low-restraint group.

Care must be taken when drawing inferences from studies linking the TFEQ-R to lower caloric intake and/or healthier intake (e.g., fewer fats/sweets, more fruits and vegetables) as restrained

eaters are known to underestimate their caloric intake to a greater degree than restrained eaters (for a review, see Maurer et al., 2006). Furthermore, the source of the underreporting seems to be disproportionately accounted for by the unhealthiest foods (Maurer et al., 2006).

Eating Disorders and Psychopathology

A plethora of studies report cross-sectional correlations for the TFEQ-R and measures of eating disorders. For example, Boerner et al. (2004) found a correlation of .43 for men and .52 for women between the TFEQ-R and the Bulimia Test–Revised (BULIT-R; Thelen et al., 1991). Similarly, the correlation between the TFEQ-R and the EAT (Garner & Garfinkel, 1979) was .45 for men and .64 for women. Atlas et al. (2002) found that the TFEQ-R correlated with the BULIT-R at .47 for White and .69 for African American college women. Ricciardelli, Tate, and Williams (1997) found evidence that body dissatisfaction may mediate the relationship between the TFEQ-R and the BULIT-R. However, their conclusions are limited by the cross-sectional nature of their research design. Rigid and flexible (Westenhoefer, 1991) control over eating appears to be differentially related to measures of eating disorders. In a field survey of 1,838 West Germans, rigid control was associated with more frequent and more severe binge episodes, whereas flexible control was associated with the opposite (Westenhoefer et al., 1999). The TFEQ-R as a whole was not related to binge frequency or severity. High scores on the TFEQ-R were associated with greater risk for using purging behaviors such as diuretics, laxatives, appetite suppressants, vomiting, physical exercise, and bodybuilding. Higher rigid control was associated with a higher risk of using all of these purging techniques except physical exercise and bodybuilding. Higher flexible control was associated with a lower risk of using diuretics or appetite suppressants and a higher likelihood of using physical exercise or bodybuilding as methods of weight control. Despite the correlations between the TFEQ-R and measures of eating-disordered attitudes and behaviors, Safer, Agras, Lowe, and Bryson (2004) reported that TFEQ-R scores did not decrease significantly during cognitive-behavioral therapy for bulimia in a sample of 134 women.

Correlations between the TFEQ-R and measures of eating-disordered symptomatology should not be interpreted as supporting a causal link between this measure of dieting and eating disorders. For one, the studies finding such a relationship were all done with nonclinical populations, and only a very small percentage was likely to suffer from an eating disorder. Also, if there were a causal link, then those bulimic individuals who are actually dieting to try to lose weight should show particularly high levels of binge eating. Instead, strict dieting is associated with reduced, rather than enhanced, binge-eating frequency (Lowe et al., 1998, 2007). Finally, the observation that TFEQ-R scores do not decrease during treatment (Safer et al., 2004) for bulimia seems to suggest that the construct of dieing tapped by the TFEQ-R is not an important factor in the maintenance of this eating disorder. Prospective studies are needed to determine what, if any, role this type of dieting may play in the development and maintenance of disordered eating.

Susceptibility to Response Sets

Allison et al. (1992) found weak correlations between the TFEQ-R and the Edwards and Marlowe-Crowne Social Desirability Scales (r = .05 and –.21, respectively). Furthermore, ratings of the social desirability of each item did not correlate with the frequency with which they were

endorsed. Finally, instructions to "fake good" and "fake bad" did not result in significantly different means on the TFEQ-R. On the basis of these results, the authors conclude that the TFEQ-R is not unduly influenced by socially desirable responding.

Predictions of Laboratory Behavior

One of the most well-known qualities of Herman and Polivy's Restraint Scale is its ability to predict disinhibited eating in the laboratory setting. In contrast, the TFEQ-R is not typically linked to disinhibited eating in preload/taste test studies (Lowe & Maycock, 1988; Rogers & Hill, 1989; Tuschl, Laessle, Platte, & Pirke, 1990; Westerterp, Nicolson, Boots, Mordant, & Westerterp, 1988; Westerterp-Plantenga et al., 1991). It is more common to find that a tendency toward disinhibited eating, as measured by the Disinhibition subscale of the TFEQ, for example, is a better predictor of overeating (e.g., Ouwens et al., 2003; van Strien et al., 2000). The discrepancy between the RS and the TFEQ-R in the prediction of disinhibited eating often has been explained by the assertion that the TFEQ-R tends to select a broad range of dieters, including those who are successful and unsuccessful, whereas the RS tends to select primarily failed dieters who have a tendency toward overeating (for a review, see van Strien, 1999). Thus, van Strien (1999) recommended that the TFEQ-R be used in conjunction with the TFEQ Disinhibition subscale to independently study the individual and combined associations of these constructs with eating behavior.

Readability

The reading level of the TFEQ-R has been estimated to be between the sixth and ninth grades (Allison & Franklin, 1993).

Availability

The TFEQ can be purchased from Harcourt Assessment (harcourtassessment.com).

Dutch Eating Behavior Questionnaire

Description

van Strien, Frijters, Bergers, et al. (1986) created the Dutch Eating Behavior Questionnaire (DEBQ) to facilitate research on the development and maintenance of human obesity. The measure was created partly in response to psychosomatic theory, externality theory, and Herman and Polivy's restraint theory, all of which suggest that obesity is attributable to overeating.

The DEBQ was created in response to the same criticisms of the RS that led Stunkard and Messick to develop the TFEQ. While the TFEQ was published before the DEBQ, the two measures were under development at about the same time. In fact, both scales borrowed items from Pudel's et al.'s (1975) Latent Obesity Questionnaire, which may partially explain any correlation observed between the DEBQ and TFEQ restraint scales. In addition to a restraint subscale that

was intended to be distinct from measures of overeating and independent of obesity status, the DEBQ includes subscales for emotional eating and external eating. The restraint subscale includes items pertaining to deliberate, planned weight control. The emotional eating subscale prompts individuals to indicate how often they experience a desire to eat as a result of unpleasant emotions such as anxiety, sadness, and boredom. The external eating subscale has items that refer to increased consumption or desire for food in the presence of food-related stimuli.

During the initial measure development process of the DEBQ, a pool of 100 items taken from previous measures, including the Eating Patterns Questionnaire and the Eating Behavior Inventory (O'Neil et al., 1979), were administered to a sample of 140 participants, including normal-weight and obese individuals. A series of factor analyses and item analyses were used to identify items that appeared factorially simple (i.e., tended to load only on one factor). In addition, some items were revised, and new items created, to increase the distinctiveness of the subscales.

The final scale consisted of 33 items divided among three subscales. The response options for each item are on a Likert-type scale with the following categories: *never* (1), *seldom* (2), *sometimes* (3), *often* (4), and *very often* (5). The subscales of the DEBQ are typically scored by calculating the average response for all items in each scale. Although the developers' intention was to create a measure with three distinct factors, a fourth factor was identified during the final analyses that represented emotional eating while bored. This fourth factor was not included as a formal subscale, as it contained items that loaded highly on other subscales, and was not of specific theoretical interest. For our purposes, all further discussion will be limited to the restraint subscale (DEBQ-R) of the DEBQ.

Sample

The final form of the DEBQ was tested on a sample of 517 male and 653 female participants, 114 of whom were obese.

Norms

Table 5.5 presents norms for the DEBQ restraint scale. Women appear to score higher on the DEBQ-R than men, and obese individuals seem to have higher scores than persons of normal weight. Care should be taken when classifying individuals as restrained and unrestrained as no empirically validated cutoff exists, and the distribution of scores varies by nationality. Although sample medians are often used to create two restraint groups, it is generally preferable to treat the DEBQ-R score (or any restraint score) as continuous when possible.

Reliability

Internal Consistency

The rigorous development process of the DEBQ resulted in a restraint factor with high internal consistency. As can be seen in Table 5.2, Cronbach's alpha is generally greater than .90. Furthermore, the scale appears to be equally reliable in normal-weight and obese individuals.

Table 5.5 Mean Dutch Eating Behavior Questionnaire–Restraint Subscale (DEBQ-R) Scores Reported in the Literature

Participants	*Author*	*n*	*Mean*	*SD*
Dutch adults	van Strien, Frijters, Bergers, and Defares (1986)	1169	2.2	0.9
Men only		498	1.8	0.8
Obese men only		71	2.3	0.8
Nonobese men only		427	1.8	0.7
Women only		642	2.5	0.9
Obese women only		73	3.0	0.8
Nonobese women only		569	2.4	0.9
Dutch college students	Ouwens, van Strien, and van der Staak (2003)	209	2.6	0.9
Dutch college students	van Strien, Cleven, and Schippers (2000)	200	2.6	0.8
Normal-weight female Dutch college students	van Strien, Herman, Engels, Larsen, and van Leeuwe (2007)	349	2.6	0.8
Overweight, nonobese Dutch women	van Strien et al. (2007)	409	3.2	0.7
American college students	Allison, Kalinsky, and Gorman (1992)	901	2.9	1.0
Men only		281	2.3	0.9
Obese men only		7	3.1	0.8
Nonobese men only		274	2.3	0.0
Women only		607	3.1	1.0
Obese women only		23	3.2	0.8
Nonobese women only		584	3.1	1.0
Australian Grade 9 female adolescents	Banasiak, Wertheim, Koerner, and Voudouris (2001)	393	2.7	0.8
English men	Wardle (1986)	45	1.9	0.8
English women	Wardle (1986)	102	2.7	0.0
German women	Laessle, Tuschl, Kotthaus, and Pirke (1989)	60	2.4	0.6

Test-Retest

In a sample of 165 adolescent girls, Banasiak, Wertheim, Koerner, and Voudouris (2001) found the test-retest reliability of the DEBQ-R to be .85 after a delay of 4 to 5 weeks. The retest coefficient for a 2-week span was .92.

Validity

Factorial Composition

Few published studies have tested the factor structure of the DEBQ. Of those that have, the majority found that a simple three-factor solution including all 33 items fits the data quite well, with a restraint factor that is clearly separate from the factors representing emotional eating and external eating (van Strien, Frijters, Bergers, et al.,1986; Wardle, 1987). van Strien, Frijters, Bergers, et al. (1986) also found the factor structure to be invariant for both genders and persons of obese and normal weight.

Two other studies investigated the factor structure of the 10-item DEBQ restraint scale. After completing both exploratory and confirmatory factor analyses, Allison et al. (1992) concluded that the DEBQ-R was best described by a unifactorial solution accounting for 68% of the variance. However, Ogden (1993) observed that the DEBQ-R contains two potentially confounded aspects of dietary restraint: attempts at food restriction and actual restrictive behavior. To examine this possibility, she conducted an exploratory factor analysis of a modified DEBQ-R in which extra items were added to questions, including the word *try,* that specifically distinguished between intended restraint and successful restraint. In addition, two new items were added: "Do you attempt to diet in order to lose weight?" and "Do you regard yourself as a successful dieter?" All items but Item 4 loaded on a single factor containing the two additional items, which suggests that individuals do not distinguish between attempts at restraint and actual restraint behaviors. However, the possibility remains that restrained eating varies on a single continuum ranging from intentions to diet to actual restrictive behaviors, as was found in studies of the TFEQ-R by Gorman et al. (1993) and Allison et al. (1992).

Factor Stability

During the measure development process, van Strien, Frijters, Bergers, et al. (1986) noted that the pattern of item-total scale correlations was similar for obese and normal-weight participants. Allison et al. (1992) conducted separate factor analyses of the DEBQ-R for obese and normal-weight participants, for men and women, and for random splits of the sample. They found that Tucker's congruence coefficients were at least .990 in each split. Based on these data, the factor stability of the DEBQ-R seems excellent.

Construct Validity: Convergent and Discriminant Validity

Preliminary evidence suggests that the restraint subscale of the DEBQ is minimally related to the other two DEBQ subscales. van Strien, Frijters, Bergers, et al. (1986) report that the DEBQ restraint scale correlated at .37 with the DEBQ emotional eating scale and .16 with the DEBQ external eating subscale in a mixed sample of normal-weight and obese individuals.

Weight and Obesity Status

The mean DEBQ-R scores of 76 friendship cliques consisting of 523 adolescent girls were correlated with mean clique BMI (r = .38; Paxton, Schutz, Wertheim, & Muir, 1999). In a randomized controlled trial of behavioral weight loss interventions, DEBQ-R scores increased significantly in the three treatment conditions but not in a control condition (Williamson et al., 2007). The sample consisted of 46 overweight (25 < BMI < 30 kg/m^2) individuals.

Ogden (1993) studied a sample of "successful," "reasonable," and "failed" dieters, who were categorized based on whether they rated their success at dieting as higher, equivalent, or lower than their attempts at dieting, respectively. DEBQ-R scores were highest among the failed dieters, lowest among the successful dieters, and intermediate among the reasonable dieters. While this finding suggests that the DEBQ-R is related to unsuccessful attempts at dieting, care must be taken when interpreting the results, as there is no assessment of the reliability or validity of the self-reported measures of dieting frequency or success used in this study.

Naturalistic Food Consumption

Several studies have reported moderate negative correlations between caloric intake and the DEBQ-R. In a sample of 50 female undergraduates and university staff, Wardle and Beales (1987) found a correlation of –.28 between the DEBQ-R and caloric intake over a 1-day period, as assessed by interviewers trained in conducting 24-hour food recalls. Similarly, in a sample of 110 Dutch women, van Strien, Frijters, Staveren, Defares, and Deurenberg (1986) reported a correlation of –.47 between the DEBQ-R and a measure of deviation from required energy intake, which was computed by subtracting the mean caloric intake across three 24-hour food recalls from an estimate of the number of calories needed for weight maintenance. This finding suggests that individuals who score high on the DEBQ-R consume fewer calories than what is needed to sustain their current body weight. Some of this difference may be the result of ingesting fewer high-calorie foods, as the DEBQ-R also correlated at –.28 with fat intake and –.38 with sugar intake. Laessle et al. (1989) also found that the DEBQ-R correlated at –.49 with a measure of caloric intake based on computer-assisted analysis of 7-day food diaries that were completed by 60 normal-weight women. Collectively, these studies seem to indicate that the DEBQ-R identifies individuals with comparatively lower food intake, which may result in negative energy balance. However, this conclusion is qualified by previously mentioned research that finds restrained eaters systematically underreport their food intake to a greater degree than unrestrained eaters and that the source of the underreporting is disproportionately accounted for by the unhealthiest foods (Stice et al., 2004, 2007).

Prediction of Laboratory Behavior

Unlike the Restraint Scale, higher scores on the DEBQ-R are not typically associated with disinhibited eating behavior in preload studies. Of the studies that failed to detect disinhibited eating following a preload, two studies found a small but significant positive relationship between the DEBQ-R and food consumption during the "taste test" (van Strien et al., 2000; Wardle & Beales, 1987), while one other did not (Ouwens et al., 2003). Despite the lack of a disinhibition effect, participants scoring high on the DEBQ-R have been known to exhibit increased food consumption following a cognitive task (Lattimore & Caswell, 2004; Wallis & Hetherington, 2004) and a task involving ego threat (Wallis & Hetherington, 2004). In addition, female

restrained eaters (as identified by a median split of DEBQ-R scores) tended to consume more calories than unrestrained eaters, when given ad libitum access to large amounts of palatable food (Jansen, 1996). Notably, unrestrained eaters were able to estimate their caloric intake quite well, while restrained eaters underestimated their intake.

Disordered Eating and Psychopathology

Like other measures of dietary restraint, the DEBQ-R is often correlated with eating-disordered attitudes and behaviors, as well as general measures of psychopathology. In a sample of 123 young adults, DEBQ-R was significantly associated with a measure of anxiety, but only for women (Jeffery & French, 1999). DEBQ-R was not associated with depression in either gender. Paxton et al. (1999) studied restraint and disordered eating in 79 friendship cliques consisting of 523 adolescent girls. The DEBQ-R was significantly correlated with mean clique scores for body image concerns and extreme weight loss behavior but not depression, self-esteem, or anxiety. Stice et al. (1997) reported correlations of .62, .53, and .69 between the DEBQ-R and the BULIT-R total score, the BULIT-R binge control subscale, and the bulimia factor of the EAT, respectively, among 117 female college students. However, some of the relationship between the DEBQ-R and measures of psychopathology may be explained by other variables. For example, in a study of 1,177 adolescent girls over a 1-year period, F. Johnson and Wardle (2005) found that the cross-sectional and prospective relationships between the DEBQ-R and symptoms of bulimia, low self-esteem, and depression were better accounted for by body dissatisfaction. The presence and later development of abnormal eating attitudes was the only outcome with which restraint was independently associated.

Susceptibility to Response Sets

The DEBQ-R does not appear to be unduly influenced by social desirability responding or dissimulation. The correlation between the DEBQ-R and social desirability scales such as the Marlowe-Crowne Social Desirability Scale ($r = -.08$) and the Edwards Social Desirability Scale ($r = -.24$) appears to be weak and statistically nonsignificant (correlation coefficients from Allison et al., 1992; also see Corrigan & Ekstrand, 1988; van Strien, Frijters, Roosen, Knuiman-Hijl, & Defares, 1985). When each item of the DEBQ-R was rated for its social desirability, Allison et al. (1992) found that the social desirability ratings correlated with item endorsement at .67, indicating that the more desirable items were endorsed more frequently. When participants were instructed to "fake good" or "fake bad," the resulting mean DEBQ-R scores were not significantly lower or higher than when such instructions were not given. These findings indicate that the DEBQ-R scale has good discriminant validity.

Readability

The reading level of the TFEQ-R has been estimated to be between the fifth and eighth grades (Allison & Franklin, 1993).

Availability

The DEBQ-R was originally printed in van Strien, Frijters, Bergers, et al. (1986).

Other Scales

The RS, TFEQ-R, and DEBQ-R are typically the measures of choice when studying restrained eating. However, there are a few other scales worth mentioning, although most of the following lack much psychometric evidence to support their reliability or validity. One exception is the restraint scale of the Eating Disorders Examination, which is available in questionnaire (EDE-Q; Fairburn & Beglin, 1994) and interview (EDE; Fairburn & Cooper, 1993) forms. The EDE is primarily a diagnostic tool for anorexia and bulimia nervosa, for which its reliability and validity have been well demonstrated. However, the EDE is intended for use only in eating-disordered samples. As such, the restraint subscale is not appropriate for use with nonclinical samples.

A restraint interview was created by Rand and Kuldau (1991) for use with nonclinical samples that may have certain advantages, including the potential for phone-based assessment and no requirement of reading skills on the part of the subject. Also, there is some thought that interviews may be less susceptible to dissimulation, given a skilled interviewer. Child versions of the RS and TFEQ were developed by Hill et al. (Hill, Rogers, & Blundell, 1989; Hill, Weaver, & Blundell, 1990). Other instruments that purport to measure restraint have been developed by Coker and Roger (1990) and Smead (1990).

Relationships Among the Restraint Scales

Intercorrelations among the RS, TFEQ-R, and DEBQ-R are illustrated in Table 5.6. The TFEQ-R borrowed items from the RS, and the TFEQ-R and DEBQ-R both contain items from Pudel et al.'s (1975) Latent Obesity Scale. Thus, high correlations among the three restraint measures are not surprising. The correlations among measures appear to be similar for men and women. In contrast, correlations among scales appear to be lower for overweight than normal-weight individuals, especially for correlations between the RS and the other two scales. van Strien et al. (2007) report that the correlations between the RS (including the total score and both subscales) and the DEBQ-R and TFEQ-R are significantly lower for overweight women than normal-weight women. As discussed previously, the RS was not designed for use with overweight individuals and has questionable validity when used with this population. The lower correlations among restraint scales for overweight individuals are further evidence that the restraint constructs applied to normal-weight individuals do not translate perfectly to overweight samples.

Wardle (1986) reported greater correlations between the DEBQ-R and the RS CD subscale (r = .75 for women and r = .76 for men) than the RS WF subscale (r = .24 for women and r = .37 for men). A similar pattern of results was reported by Boerner et al. (2004) for the relationship between the TFEQ-R and RS subscales, as well as by van Strien et al. (2007) for the RS subscales and both the TFEQ-R and DEBQ-R. These findings suggest that the three scales share common variance related to cognitive restraint but that the RS WF subscale measures a dimension that the other two scales do not address.

The relationships among the three restraint measures have also been tested by conducting factor analyses on the scale scores for the restraint scales and sometimes other measures of eating behavior and weight concerns. For example, Allison et al. (1992) took the factors identified in factor analyses of each individual restraint scale and performed a second-order principal components factor analysis on these factors to look for overlap among the scales. The result was a three-factor solution. The first factor represented cognitive restraint and had high loadings from

Table 5.6 Intercorrelations Among Restraint Scales

Reference	*n*	*Coefficient alpha*	*Sample Characteristics*
RS and TFEQ-R			
Allison, Kalinsky, and Gorman (1992)	901	.74	Obese and normal-weight college students
Boerner, Spillane, Anderson, and Smith (2004)	214	.63	Male college students
Boerner et al. (2004)	215	.68	Female college students
Laessle, Tuschl, Kotthaus, and Pirke (1989)	60	.35	Normal-weight women
Ouwens, van Strien, and van der Staak (2003)	209	.73	Female college students
van Strien, Cleven, and Schippers (2000)	200	.57	Female college students
Williamson et al. (2007)	46	.51	Overweight men and women
van Strien, Herman, Engels, Larsen, and van Leeuwe (2007)	349	.74	Normal-weight female college students
van Strien et al. (2007)	409	.35	Overweight, nonobese women
RS and DEBQ-R			
Allison et al. (1992)	901	.80	Obese and normal-weight college students
Laessle et al. (1989)	60	.59	Normal-weight women
Ouwens et at. (2003)	209	.69	Female college students
Stice, Ozer, and Kees (1997)	117	.83	Female college students
van Strien et al. (2000)	200	.55	Female college students
Williamson et al. (2007)	46	.55	Overweight men and women
van Strien et al. (2007)	349	.71	Normal-weight female college students
van Strien et al. (2007)	409	.36	Overweight, nonobese women
Wardle (1986)	147	.72	Female college students
Wardle (1986)	147	.75	Male college students
TFEQ and DEBQ-R			
Allison et al. (1992)	901	.89	Obese and normal-weight college students
Laessle et al. (1989)	60	.66	Normal-weight women
Ouwens et at. (2003)	209	.85	Female college students
van Strien et al. (2000)	200	.75	Female college students
Williamson et al. (2007)	46	.69	Overweight men and women
van Strien et al. (2007)	349	.86	Normal-weight female college students
van Strien et al. (2007)	409	.66	Overweight, nonobese women

NOTE: DEBQ-R = Dutch Eating Behavior Questionnaire–Restraint subscale; RS = Restraint Scale; TFEQ-R = Three-Factor Eating Questionnaire–Restraint scale.

the RS Concern With Dieting subscale, the DEBQ-R, and the TFEQ-R Factor I (cognitive restraint). The second factor consisted of Factor I (cognitive restraint) and Factor II (behavioral restraint) from the TFEQ-R and was therefore determined to be representative of a general restraint factor specific to the TFEQ-R. The third factor included only the RS Weight Fluctuation subscale. The authors concluded that the three scales share some common variance but that the TFEQ-R is the only scale that measures behavioral restraint, and the RS WF subscale is the only measure of weight fluctuation.

Laessle et al. (1989) conducted a factor-analytic investigation that included the RS, the DEBQ-R, and the TFEQ restraint and disinhibition scales, as well as measures of weight history, self-reported mean daily caloric intake, disordered eating, and body figure consciousness. The first factor had high loadings from the RS, scales representing counterregulatory or disinhibited eating (the Eating Disorder Inventory Bulimia subscale and the TFEQ Disinhibition subscale), and measures representing body concern (Eating Disorder Inventory Body Dissatisfaction and Drive for Thinness subscales and the Body Shape Questionnaire). The second factor had high loadings from the RS and weight-related measures (BMI, maximum BMI, and a BMI fluctuation index). The third factor had high loadings from the TFEQ-R and the DEBQ-R, as well as a negative loading on mean caloric intake.

van Strien et al. (2007), noting the three-factor solution obtained by Laessle et al. (1989), conducted a series of confirmatory factor analyses to determine how the three measures of restraint would load on three factors representing overeating, dieting, and body dissatisfaction. The overeating factor included the Eating Disorder Inventor Bulimia subscale, the DEBQ Emotional Eating and External Eating subscales, and the question, "Have you ever had an eating binge, i.e., you ate an amount of food others would consider unusually large?" The dieting factor included the question, "Are you currently dieting?" The body dissatisfaction factor included the Eating Disorder Inventory Drive for Thinness and Body Dissatisfaction subscales. The best-fit models for the TFEQ-R and DEBQ-R were the ones in which these scales loaded only on the dieting factor but not the overeating or body dissatisfaction factors. This was true for normal-weight and overweight subsamples. In contrast, the best-fit model for the RS was the one in which it loaded on all three factors, rather than just the dieting factor. There was an association between dieting and overeating in the normal-weight sample that was absent in the overweight sample.

The results of Laessle et al. (1989) and van Strien et al. (2007) seem to confirm that the TFEQ-R and DEBQ-R are "purer" measures of restraint, whereas the RS taps constructs related to unsuccessful dieting such as overeating and weight fluctuation. Furthermore, the findings of van Strien et al. may explain why overweight individuals do not show disinhibited eating in preload studies; they lack the association between restraint and overeating that is present among normal-weight individuals.

Future Research Directions

One priority for future research is improving our understanding of what the different restraint scales are actually measuring. Our review makes it clear that the RS reflects both the tendency to lose control over eating and the effort to resist that tendency. The fact that the RS taps both tendencies simultaneously might be advantageous for some research questions, but the field's understanding of factors that promote and inhibit overeating would be better served by research designs that analytically separate these two factors. Research that has categorized participants

on both the TFEQ disinhibition scale and the TFEQ restraint scale (e.g., Westenhoefer et al., 1994) represents one way of doing this.

For the TFEQ-R and the DEBQ-R, it is becoming apparent that these scales do not identify individuals who are in negative energy balance or who are restricting their energy intake relative to unrestrained eaters (Stice et al., 2004, 2007). However, they may be restricting their intake relative to what they would *like* to eat (Lowe & Butryn, 2007). Although forced preloads do not elicit counterregulatory eating in restrained eaters identified by these scales, it is possible that such individuals would nonetheless show poorer eating regulation in situations in which multiple disinhibiting influences are operating simultaneously (e.g., a social gathering where alcohol and a variety of palatable foods are being consumed). It is possible that simply providing ice cream following a milkshake preload simply does not constitute a disinhibiting context powerful enough to overcome these restrained eaters' efforts to avoid overconsumption (e.g., Jansen, 1996).

Another major implication of the evidence reviewed in this chapter is that none of the measures of restrained eating reflects *dieting* as that term is usually understood—that is, losing weight by eating less than needed. Indeed, as Lowe (1993) suggested, "dieting to lose weight" and "restrained eating" appear to be two different constructs that are associated with different and sometimes opposing effects on behavior. Although measures of restrained eating have been shown to be related to a variety of domains (affective, cognitive, behavioral, physiological, and genetic), it cannot be assumed that these associations are due to hypocaloric dieting. Thus, future research is needed to study the effects of "restrained eating" separately from dieting (both in terms of self-labeled current dieting and documented weight loss dieting). Furthermore, if Lowe and Levine (2005) are correct that most restrained eating research should be interpreted in terms of the consequences of eating less than desired rather than eating less than needed, then new explanations may be needed for many of the findings documented in the restraint literature.

Finally, it is very important to keep in mind that the vast majority of research on restrained eating has been correlational in nature. This, of course, leaves open the question of whether restraint plays the causal role it is assumed to play in eating disregulation and eating disorders. Indeed, when dieting status has been experimentally manipulated, its effects are often opposite (e.g., Foster, Wadden, Kendall, Stunkard, & Vogt, 1996; Presnell & Stice, 2003) to those predicted by the original restraint model (Herman & Polivy, 1975, 1984). This suggests that restrained eating per se may not be responsible for the effects that are often associated with it. Alternatively, since most normal-weight restrained eaters are prone toward weight gain, it may be that restraint acts to moderate a predisposition toward weight gain such that restraint slows but usually does not prevent eventual weight gain. Also, it is important to keep in mind that, to the extent that restrained eating does have causal effects on behavior, they may be quite different depending on why a person is attempting to exercise dietary restraint. For example, an anorexic restrictor, a normal-weight person who is struggling to avoid weight gain, and an obese binge eater may all be "restrained eaters," but the form and consequences of such restraint may be quite different in each.

References

Allison, D. B., & Franklin, R. D. (1993). The readability of three measures of dietary restraint. *Psychotherapy in Private Practice, 12*(3), 53–57.

Allison, D. B., Gorman, B. S., & Primavera, L. H. (1993). Some of the most common questions asked of statistical consultants: Our favorite responses and recommended readings. *Genetic, Social, and General Psychology Monographs, 119,* 153–185.

Allison, D. B., Kalinsky, L. B., & Gorman, B. S. (1992). A comparison of the psychometric properties of three measures of dietary restraint. *Psychological Assessment, 4,* 391–398.

Atlas, J., Smith, G., Hohlstein, L., McCarthy, D., & Kroll, L. (2002). Similarities and differences between Caucasian and African American college women on eating and dieting expectancies, bulimic symptoms, dietary restraint, and disinhibition. *International Journal of Eating Disorders, 32,* 326–334.

Banasiak, S., Wertheim, E., Koerner, J., & Voudouris, N. (2001). Test-retest reliability and internal consistency of a variety of measures of dietary restraint and body concerns in a sample of adolescent girls. *International Journal of Eating Disorders, 29,* 85–89.

Bandini, L. G., Schoeller, D. A., Dyr, H. N., & Dietz, W. H. (1990). Validity of reported energy intake in obese and nonobese adolescents. *American Journal of Clinical Nutrition, 52,* 421–425.

Beiseigel, J., & Nickols-Richardson, S. (2004). Cognitive eating restraint scores are associated with body fatness but not with other measures of dieting in women. *Appetite, 43,* 47–53.

Bellisle, F., Clement, K., LeBarzic, M., LeGall, A., GuyGrand, B., & Basdevant, A. (2004). The eating inventory and body adiposity from leanness to massive obesity: A study of 2509 adults. *Obesity Research, 12,* 2023–2030.

Birch, L. L., Fisher, J. O., & Davison, K. K. (2003). Learning to overeat: Maternal use of restrictive practices promotes girls' eating in the absence of hunger. *American Journal of Clinical Nutrition, 78,* 215–220.

Blanchard, F. A., & Frost, R. O. (1983). Two factors of restraint: Concern for dieting and weight fluctuation. *Behaviour Research and Therapy, 21,* 259–267.

Block, G., Hartman, A. M., Dresser, C. M., Carroll, M. D., Gannon, J., & Gardner, L. (1986). A data-based approach to diet questionnaire design and testing. *American Journal of Epidemiology, 124,* 453–469.

Boerner, L. M., Spillane, N. S., Anderson, K. G., & Smith, G. T. (2004). Similarities and differences between women and men on eating disorder risk factors and symptom measures. *Eating Behaviors, 5,* 209–222.

Bond, M. J., McDowell, A. J., & Wilkinson, J. Y. (2001). The measurement of dietary restraint, disinhibition and hunger: An examination of the factor structure of the three factor eating questionnaire (TFEQ). *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity, 25,* 900–906.

Bourne, S. K., Bryant, R. A., Griffiths, R. A., Touyz, S. W., & Beumont, P. J. V. (1998). Bulimia nervosa, restrained, and unrestrained eaters: A comparison of non-binge eating behavior. *International Journal of Eating Disorders, 24,* 185–192.

Butryn, M. L., & Wadden, T. A. (2005). Treatment of overweight in children and adolescents: Does dieting increase the risk of eating disorders? *International Journal of Eating Disorders, 37,* 285–293.

Chernyak, Y., & Lowe, M. R. (2007, October). *Differentiating drive for thinness and drive to be thin: Restrained eaters and bulimic individuals have different motives for dieting.* Poster session presented at the annual meeting of the Eating Disorders Research Society, Pittsburg, PA.

Coker, S., & Roger, D. (1990). The construction and preliminary validation of a scale for measuring eating disorders. *Journal of Psychosomatic Research, 34,* 223–231.

Cole, S., & Edelmann, R. (1987). Restraint, eating disorders and need to achieve in state and public school subjects. *Personality and Individual Differences, 8,* 475–482.

Corrigan, S. A., & Ekstrand, M. L. (1988). An investigation of the construct validity of the Dutch Restrained Eating Scale. *Addictive Behaviors, 13,* 303–306.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart & Wilson.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York: John Wiley.

de Castro, J. (1995). The relationship of cognitive restraint to the spontaneous food and fluid intake of free-living humans. *Physiology & Behavior, 57,* 287–295.

De Lauzon-Guillain, B., Basdevant, A., Romon, M., Karlsson, J., Borys, J. M., Charles, M. A., et al. (2006). Is restrained eating a risk factor for weight gain in a general population? *American Journal of Clinical Nutrition, 83,* 132–138.

Drewnowski, A., Riskey, D., & Desor, J. A. (1982). Feeling fat yet unconcerned: Self-reported overweight and the restraint scale. *Appetite, 3,* 273–279.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16,* 363–370.

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

Ferguson, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika, 6,* 323–329.

Ferguson, K. J., Brink, P. J., Wood, M., & Koop, P. M. (1992). Characteristics of successful dieters as measured by guided interview responses and restraint scale scores. *Journal of the American Dietetic Association, 92,* 1119–1121.

Foster, G. D., Wadden, T. A., Kendall, P. C., Stunkard, A. J., & Vogt, R. A. (1996) Psychological effects of weight loss and regain: A prospective evaluation. *Journal of Consulting and Clinical Psychology, 64,* 752–757.

Foster, G. D., Wadden, T. A., Swain, R. M., Stunkard, A. J., Platte, P., & Vogt, R. A. (1998). The eating inventory in obese women: Clinical correlates and relationship to weight loss. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity, 22,* 778–785.

French, S. A., Jeffery, R. W., & Wing, R. R. (1994). Food intake and physical activity: A comparison of three measures of dieting. *Addictive Behaviors, 19,* 401–409.

Ganley, R. M. (1988). Emotional eating and how it relates to dietary restraint, disinhibition, and perceived hunger. *International Journal of Eating Disorders, 7,* 635–647.

Garner, D. M., & Fairburn, C. G. (1988). Relationship between anorexia nervosa and bulimia nervosa: Diagnostic implications. In D. M. Garner & P. E. Garfinkel (Eds.), *Diagnostic issues in anorexia nervosa and bulimia nervosa* (p. 56). New York: Brunner/Mazel.

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

Gibson, W. A. (1967). A latent structure for the simplex. *Psychometrika, 32,* 33–46.

Gorman, B. S., Allison, D. B., & Primavera, L. H. (1993). *The scalability of items of the three-factor eating questionnaire restraint scale: When is a personality scale a "scale"?* Arlington, VA: Eastern Physiological Association Convention.

Griffiths, R. A., MalliaBlanco, R., Boesenberg, E., Ellis, C., Fischer, K., Taylor, M., et al. (2000). Restrained eating and sociocultural attitudes to appearance and general dissatisfaction. *European Eating Disorders Review, 8,* 394–402.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hawkins, R. C. I., & Clement, P. F. (1984). Binge eating: Measurement problems and a conceptual model. In R. C. Hawkins, W. J. Fremouw, & P. F. Clement (Eds.), *The binge purge syndrome: Diagnosis, treatment, and research* (pp. 229–253). New York: Springer.

Hays, N., Bathalon, G., Roubenoff, R., McCrory, M., & Roberts, S. (2006). Eating behavior and weight change in healthy postmenopausal women: Results of a 4-year longitudinal study. *Journals of Gerontology: Series A: Biological Sciences and Medical Sciences, 61A,* 608–615.

Heatherton, T. F., Herman, C. P., & Polivy, J. (1991). Effects of physical threat and ego threat on eating behavior. *Journal of Personality and Social Psychology, 60,* 138–143.

Heatherton, T. F., Herman, C. P., Polivy, J., King, G. A., & McGree, S. T. (1988). The (mis)measurement of restraint: An analysis of conceptual and psychometric issues. *Journal of Abnormal Psychology, 97,* 19–28.

Heatherton, T. F., & Polivy, J. (1992). *Chronic dieting and eating disorders: A spiral model.* In C. H. Janis, D. L. Tennenbaum, S. E. Hobfoll, & M. A. P. Stephens (Eds.), *The etiology of bulimia nervosa: The individual and familial context* (pp. 133–155). Washington, DC: Hemisphere Publishing Corporation.

Heatherton, T. F., Polivy, J., & Herman, C. P. (1989). Restraint and internal responsiveness: Effects of placebo manipulations of hunger state on eating. *Journal of Abnormal Psychology, 98,* 89–92.

Heatherton, T. F., Polivy, J., & Herman, C. P. (1991). Restraint, weight loss, and variability of body weight. *Journal of Abnormal Psychology, 100,* 78–83.

Herman, C. P., & Mack, D. (1975). Restrained and unrestrained eating. *Journal of Personality, 43,* 647–660.

Herman, C. P., & Polivy, J. (1975). Anxiety, restraint, and eating behavior. *Journal of Abnormal Psychology, 84,* 666–672.

Herman, C. P., & Polivy, J. (1980). *Obesity.* Philadelphia: Saunders.

Herman, C. P., & Polivy, J. (1984). A boundary model for the regulation of eating. In A. J. Stunkard & E. Stellar (Eds.), *Eating and its disorders* (pp. 141–156). New York: Raven.

Hibscher, J. A., & Herman, C. P. (1977). Obesity, dieting, and the expression of "obese" characteristics. *Journal of Comprehensive Physiological Psychology, 91,* 374–380.

Hill, A. J., Rogers, P. J., & Blundell, J. E. (1989). Dietary restraint in young adolescent girls: A functional analysis. *British Journal of Clinical Psychology, 28,* 165–176.

Hill, A. J., Weaver, C., & Blundell, J. E. (1990). Dieting concerns of 10-year-old girls and their mothers. *British Journal of Clinical Psychology, 29,* 346–348.

Horst, P. (1965). *Factor analysis of data matrices.* New York: Holt, Rinehart and Winston.

Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62,* 429–440.

Hyland, M. E., Irvine, S. H., Thacker, C., & Dan, P. L. (1989). Psychometric analysis of the Stunkard-Messick Eating Questionnaire (SMEQ) and comparison with the Dutch Eating Behavior Questionnaire (DEBQ). *Current Psychology: Research & Reviews, 8,* 228–233.

Jansen, A. (1996). How restrained eaters perceive the amount they eat. *British Journal of Clinical Psychology, 35,* 381–392.

Jeffery, R. W., & French, S. A. (1999). Preventing weight gain in adults: The pound of prevention study. *American Journal of Public Health, 89,* 747–751.

Johnson, F., & Wardle, J. (2005). Dietary restraint, body dissatisfaction, and psychological distress: A prospective analysis. *Journal of Abnormal Psychology, 114,* 119–125.

Johnson, W. G., Corrigan, S. A., Crusco, A. H., & Schlundt, D. G. (1986). Restraint among bulimic women. *Addictive Behaviors, 11,* 351–354.

Johnson, W. G., Lake, L., & Mahan, J. M. (1983). Restrained eating: Measuring an elusive construct. *Addictive Behaviors, 8,* 413–418.

Keys, A., Brozek, K., Henschel, A., Mickelsen, O., & Taylor, H. L. (1950). *The biology of human starvation.* Minneapolis: University of Minnesota Press.

Kickham, K., & Gayton, W. F. (1977). Social desirability and the restraint scale. *Psychological Reports, 40,* 550.

Killen, J. D., Hayward, C., Wilson, D. M., & Taylor, C. B. (1994). Factors associated with eating disorder symptoms in a community sample of 6th and 7th grade girls. *International Journal of Eating Disorders, 15,* 357–367.

Killen, J. D., Taylor, C. B., Hayward, C., Haydel, K. F., Wilson, D. M., Hammer, L., et al. (1996). Weight concerns influence the development of eating disorders: A 4-year prospective study. *Journal of Consulting and Clinical Psychology, 64,* 936–940.

Killen, J. D., Taylor, C. B., Hayward, C., & Wilson, D. M. (1994). Pursuit of thinness and onset of eating disorder symptoms in a community sample of adolescent girls: A three-year prospective analysis. *International Journal of Eating Disorders, 16,* 227–238.

Kleifield, E., & Lowe, M. R. (1991). Weight loss and sweetness preferences: The effects of recent versus past weight loss. *Physiology and Behavior, 49,* 1037–1042.

Klem, M. L., Klesges, R. C., Bene, C. R., & Mellon, M. W. (1990). A psychometric study of restraint: The impact of race, gender, weight and marital status. *Addictive Behaviors, 15,* 147–152.

Klem, M. L., Klesges, R. C., & Shadish, W. (1990, November). *Application of confirmatory factor analysis to the dietary restraint scale.* Paper presented to the Association for the Advancement of Behavior Therapy, San Francisco.

Klesges, R. C., Isbell, T. R., & Klesges, L. M. (1992). Relationship between dietary restraint, energy intake, physical activity, and body weight: A prospective analysis. *Journal of Abnormal Psychology, 101,* 668–674.

Klesges, R. C., Klem, M. L., Epkins, C. C., & Klesges, L. M. (1991). A longitudinal evaluation of dietary restraint and its relationship to changes in body weight. *Addictive Behaviors, 16,* 363–368.

Laessle, R. G., Tuschl, R. J., Kotthaus, B. C., & Pirke, K. M. (1989). A comparison of the validity of three scales for the assessment of dietary restraint. *Journal of Abnormal Psychology, 98,* 504–507.

Lahteenmaki, L., & Tuorila, H. (1995). Three-factor eating questionnaire and the use and liking of sweet and fat among dieters. *Physiology & Behavior, 57,* 81–88.

Larsen, J. K., van Strien, T., Eisinga, R., Herman, C. P., & Engels, R. C. M. E. (2007). Dietary restraint: Intention versus behavior to restrict food intake. *Appetite, 49,* 100–108.

Lattimore, P., & Caswell, N. (2004). Differential effects of active and passive stress on food intake in

restrained and unrestrained eaters. *Appetite, 42,* 167–173.

Lichtman, S. W., Pisarska, K., Berman, E., Pestone, M., Dowling, H., Offenbacher, E., et al. (1992). Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. *New England Journal of Medicine, 327,* 1893–1898.

Livingstone, M. B., Prentice, A. M., & Strain, J. J. (1990). Accuracy of weighed dietary records in studies of diet and health. *British Medical Journal, 300,* 708–712.

Lowe, M. R. (1984). Dietary concern, weight fluctuations and weight status: Further explorations of the restraint scale. *Behaviour Research and Therapy, 22,* 243–248.

Lowe, M. R. (1993). The effects of dieting on eating behavior: A three-factor model. *Psychological Bulletin, 114,* 100–121.

Lowe, M. R., Annunziato, R. A., Markowitz, J. T., Didie, E., Bellace, D. L., Riddell, L., et al. (2006). Multiple types of dieting prospectively predict weight gain during the freshman year of college. *Appetite, 47,* 83–90.

Lowe, M. R., & Butryn, M. (2007). Hedonic hunger: A new dimension of appetite? *Physiology & Behavior, 91,* 432–439.

Lowe, M. R., Gleaves, D. H., & Murphy-Eberenz, K. P. (1998). On the relation of dieting and bingeing in bulimia nervosa. *Journal of Abnormal Psychology, 107,* 263–271.

Lowe, M. R., & Kleifield, E. (1988). Cognitive restraint, weight suppression, and the regulation of eating. *Appetite, 10,* 159–168.

Lowe, M. R., & Kral, T. V. E. (2006). Stress-induced eating in restrained eaters may not be caused by stress or restraint. *Appetite, 46,* 16–21.

Lowe, M. R., & Levine, A. S. (2005). Eating motives and the controversy over dieting: Eating less than needed versus less than wanted. *Obesity Research, 13,* 797–805.

Lowe, M. R., & Maycock, B. (1988). Restraint, disinhibition, hunger and negative affect eating. *Addictive Behaviors, 13,* 369–377.

Lowe, M. R., Thomas, J. G., Safer, D. L., & Butryn, M. L. (2007). The relationship of weight suppression and dietary restraint to binge eating in bulimia nervosa. *International Journal of Eating Disorders, 40,* 640–644.

Lowe, M. R., Whitlow, J. W., & Bellwoar, V. (1991). Eating regulation: The role of restraint, dieting, and weight. *International Journal of Eating Disorders, 10,* 461–471.

Maurer, J., Taren, D. L., Teixeira, P. J., Thomson, C. A., Lohman, T. G., Going, S. B., et al. (2006). The psychosocial and behavioral characteristics related to energy misreporting. *Nutrition Reviews, 64*(Pt. 1), 53–66.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113,* 181–190.

Mazzeo, S., Aggen, S., Anderson, C., Tozzi, F., & Bulik, C. (2003). Investigating the structure of the eating inventory (three-factor eating questionnaire): A confirmatory approach. *International Journal of Eating Disorders, 34,* 255–264.

McCrae, R. R., & Costa, P. T. (1983a). Joint factors in self-reports and ratings: Neuroticism, extraversion and openness to experience. *Personality and Individual Differences, 4,* 245–255.

McCrae, R. R., & Costa, P. T. (1983b). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27,* 82–99.

Neale, B., Mazzeo, S., & Bulik, C. (2003). A twin study of dietary restraint, disinhibition and hunger: An examination of the eating inventory (three factor eating questionnaire). *Twin Research, 6,* 471–478.

Nisbett, R. E. (1972). Hunger, obesity, and the ventromedial hypothalamus. *Psychological Review, 79,* 433–453.

Oates-Johnson, T., & DeCourville, N. (1999). Weight preoccupation, personality, and depression in university students: An interactionist perspective. *Journal of Clinical Psychology, 55,* 1157–1166.

Ogden, J. (1993). The measurement of restraint: Confounding success and failure? *International Journal of Eating Disorders, 13,* 69–76.

O'Neil, P. M., Currey, H. S., Hirsch, A. A., Malcom, R. J., Sexauer, J. D., Riddle, F. E., et al. (1979). Development and validation of the eating behavior inventory. *Journal of Psychopathology and Behavioral Assessment, 1,* 123–132.

Ouwens, M., van Strien, T., & van der Staak, C. F. (2003). Tendency toward overeating and restraint as predictors of food consumption. *Appetite, 40,* 291–298.

Overduin, J., & Jansen, A. (1996). A new scale for use in non-clinical research into disinhibitive eating.

Personality and Individual Differences, 20, 669–677.

Paxton, S. J., Schutz, H. K., Wertheim, E. H., & Muir, S. L. (1999). Friendship clique and peer influences on body image concerns, dietary restraint, extreme weight-loss behaviors, and binge eating in adolescent girls. *Journal of Abnormal Psychology, 108,* 255–266.

Polivy, J., & Herman, C. P. (1983). *Breaking the diet habit: The natural weight alternative.* New York: Basic Books.

Polivy, J., & Herman, C. P. (1985). Dieting and binging: A causal analysis. *American Psychologist, 40,* 193–201.

Polivy, J., & Herman, C. P. (1987). Diagnosis and treatment of normal eating. *Journal of Consulting and Clinical Psychology, 55,* 635–644.

Polivy, J., Herman, C. P., & Howard, K. (1988). The restraint scale: Assessment of dieting. In M. Hersen & A. S. Bellack (Eds.), *Dictionary of behavioral assessment techniques* (p. 377). New York: Pergamon.

Prentice, A. M., Black, A. E., Coward, W. A., Davies, H. L., Goldberg, G. L., & Murgatroyd, P. (1986). High levels of energy expenditure in obese women. *British Medical Journal, 292,* 983–987.

Presnell, K., & Stice, E. (2003). An experimental test of the effect of weight-loss dieting on bulimic pathology: Tipping the scales in a different direction. *Journal of Abnormal Psychology, 112,* 166–170.

Prussin, R. A., & Harvey, P. D. (1991). Depression, dietary restraint, and binge eating in female runners. *Addictive Behaviors, 16,* 295–301.

Pudel, V., Metzdorff, M., & Oetting, M. (1975). Zur personlichkeit adiposer in psychologischen tests unter berucksichtigung latent fettsuchtiger [The personality of obese persons in psychological tests with special consideration on latent obesity]. *Zeitschrift fur Psychosomatische Medizin und Psychoanalyse, 21,* 345–361.

Rand, C. S., & Kuldau, J. M. (1991). Restrained eating (weight concerns) in the general population and among students. *International Journal of Eating Disorders, 10,* 699–708.

Ricciardelli, L., Tate, D., & Williams, R. (1997). Body dissatisfaction as a mediator of the relationship between dietary restraint and bulimic eating patterns. *Appetite, 29,* 43–54.

Ricciardelli, L., & Williams, R. (1997). A two-factor model of dietary restraint. *Journal of clinical psychology, 53,* 123–131.

Rogers, P. J., & Hill, A. J. (1989). Breakdown of dietary restraint following mere exposure to food stimuli: Interrelationships between restraint, hunger, salivation, and food intake. *Addictive Behavior, 14,* 387–397.

Rotenberg, K. J., & Flood, D. (1999). Loneliness, dysphoria, dietary restraint and eating behavior. *International Journal of Eating Disorders, 25,* 55–64.

Rotenberg, K. J., & Flood, D. (2000). Dietary restraint, attributional styles for eating, and preloading effects. *Eating Behaviors, 1,* 63–78.

Ruderman, A. J. (1983). The restraint scale: A psychometric investigation. *Behaviour Research and Therapy, 21,* 253–258.

Ruderman, A. J. (1985). Restraint, obesity and bulimia. *Behaviour Research and Therapy, 23,* 151–156.

Ruderman, A. J. (1986). Dietary restraint: A theoretical and empirical review. *Psychological Bulletin, 99,* 247–262.

Ruderman, A. J., & Grace, P. S. (1987). Restraint, bulimia, and psychopathology. *Addictive Behaviors, 12,* 249–255.

Ruderman, A. J., & Grace, P. S. (1988). Bulimics and restrained eaters: A personality comparison. *Addictive Behaviors, 13,* 359–368.

Russell, G. (1979). Bulimia nervosa: An ominous variant of anorexia nervosa. *Psychological Medicine, 9,* 429–448.

Safer, D., Agras, W. S., Lowe, M., & Bryson, S. (2004). Comparing two measures of eating restraint in bulimic women treated with cognitive-behavioral therapy. *International Journal of Eating Disorders, 36,* 83–88.

Sarwer, D. B., & Wadden, T. A. (1999). The treatment of obesity: What's new, what's recommended. *Journal of Women's Health and Gender-Based Medicine, 8,* 483–493.

Scagliusi, F. B., Polacow, V. O., Cordas, T. A., Coelho, D., Alvarenga, M., Philippi, S. T., et al. (2005). Test-retest reliability and discriminant validity of the restraint scale translated into Portuguese. *Eating Behaviors, 6,* 85–93.

Schachter, S., & Rodin, J. (1974). *Obese humans and rats.* Washington, DC: Erbaum/Halsted.

Simmons, J. R., Smith, G. T., & Hill, K. K. (2002). Validation of eating and dieting expectancy measures in two adolescent samples. *International Journal of Eating Disorders, 31,* 461–473.

Smead, V. S. (1990). A psychometric investigation of the rigorous eating scale. *Psychological Reports, 67,* 555–561.

Smith, M. C., & Thelen, M. H. (1984). Development and validation of a test for bulimia. *Journal of Consulting and Clinical Psychology, 52,* 863–872.

Stein, D. M. (1988). The scaling of restraint and the prediction of eating. *International Journal of Eating Disorders, 7,* 713–717.

Stice, E., Cooper, J., Schoeller, D., Tappe, K., & Lowe, M. (2007). Are dietary restraint scales valid measures of moderate-to long-term dietary restriction? Objective biological and behavioral data suggest not. *Psychological Assessment, 19,* 449–458.

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

Stice, E., Killen, J. D., Hayward, C., & Taylor, C. B. (1998). Age of onset for binge eating and purging during late adolescence: A 4-year survival analysis. *Journal of Abnormal Psychology, 107,* 671–675.

Stice, E., Ozer, S., & Kees, M. (1997). Relation of dietary restraint to bulimic symptomatology: The effects of the criterion confounding of the restraint scale. *Behaviour Research and Therapy, 35,* 145–152.

Stice, E., Presnell, K., Shaw, H., & Rohde, P. (2005). Psychological and behavioral risk factors for obesity onset in adolescent girls: A prospective study. *Journal of Consulting and Clinical Psychology, 73,* 195–202.

Striegel-Moore, R. H., Silberstein, L. R., & Rodin, J. (1986). Toward an understanding of risk factors for bulimia. *American Psychologist, 41,* 246–263.

Stunkard, A. J. (1981). *The body weight regulatory system: Normal and distributed mechanisms.* New York: Raven.

Stunkard, A. J., & Messick, S. (1985). The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research, 29,* 71–83.

Stunkard, A. J., & Messick, S. (1988). *The eating inventory.* San Antonio, TX: Psychological Corporation.

Thelen, M. H., Farmer, J., Wonderlich, S., & Smith, M. (1991). A revision of the bulimia test: The BULIT-R. *Psychological Assessment, 3,* 119–124.

Tholin, S., Rasmussen, F., Tynelius, P., & Karlsson, J. (2005). Genetic and environmental influences on eating behavior: The Swedish young male twins study. *American Journal of Clinical Nutrition, 81,* 564–569.

Tiggemann, M. (1994). Dietary restraint as a predictor of reported weight loss and affect. *Psychological Reports, 75*(Pt. 2), 1679–1682.

Timmerman, G. M., & Gregg, E. K. (2003). Dieting, perceived deprivation, and preoccupation with food. *Western Journal of Nursing Research, 25,* 405–418.

Tucker, L. R. (1951). *Personnel research report* (No. 984, Contract DA-49–083, Department of the Army). Princeton, NJ: ETS.

Tuschl, R. J., Laessle, R. G., Platte, P., & Pirke, K. (1990). Differences in food-choice frequencies between restrained and unrestrained eaters. *Appetite, 14,* 9–13.

Urland, G. R., & Ito, T. A. (2005). Have your cake and hate it, too: Ambivalent food attitudes are associated with dietary restraint. *Basic and Applied Social Psychology, 27,* 353–360.

van Strien, T. (1997). The concurrent validity of a classification of dieters with low versus high susceptibility toward failure of restraint. *Addictive Behaviors, 22,* 587–597.

van Strien, T. (1999). Success and failure in the measurement of restraint: Notes and data. *International Journal of Eating Disorders, 25,* 441–449.

van Strien, T., Breteler, M. H. M., & Ouwens, M. A. (2002). Restraint scale, its sub-scales concern for dieting and weight fluctuation. *Personality and Individual Differences, 33,* 791–802.

van Strien, T., Cleven, A., & Schippers, G. (2000). Restraint, tendency toward overeating and ice cream consumption. *International Journal of Eating Disorders, 28,* 333–338.

van Strien, T., Engels, R. C. M. E., van Staveren, W., & Herman, C. P. (2006). The validity of dietary restraint scales: Comment on Stice et al. (2004). *Psychological Assessment, 18,* 89–94.

van Strien, T., Frijters, J. E., Bergers, G. P. A., & Defares, P. B. (1986). Dutch Eating Behaviour Questionnaire for assessment of restrained, emotional and external eating behaviour. *International Journal of Eating Disorders, 5,* 295–315.

van Strien, T., Frijters, J. E. R., Roosen, R. G. F. M., Knuiman-Hijl, W. F. H., & Defares, P. B. (1985). Eating behavior, personality traits and body mass in women. *Addictive Behaviors, 10,* 333–343.

van Strien, T., Frijters, J. E. R., Staveren, W. A., Defares, P. B., & Deurenberg, P. (1986). The predictive validity of the Dutch restrained eating questionnaire. *International Journal of Eating Disorders, 5,* 747–755.

van Strien, T., Herman, C. P., Engels, R. C. M. E., Larsen, J. K., & van Leeuwe, J. F. J. (2007). Construct validation of the restraint scale in normal-weight and overweight females. *Appetite, 49,* 109–121.

Vartanian, L. R., Herman, C. P., & Polivy, J. (2005). Implicit and explicit attitudes toward fatness and thinness: The role of the internalization of societal standard. *Body Image, 2,* 373–381.

Wallis, D. J., & Hetherington, M. M. (2004). Stress and eating: The effects of ego-threat and cognitive demand on food intake in restrained and emotional eaters. *Appetite, 43,* 39–46.

Ward, A., & Mann, T. (2000). Don't mind if I do: Disinhibited eating under cognitive load. *Journal of Personality and Social Psychology, 78,* 753–763.

Wardle, J. (1986). The assessment of restrained eating. *Behaviour Research and Therapy, 24,* 213–215.

Wardle, J. (1987). Eating style: A validation study of the Dutch Eating Behaviour Questionnaire in normal subjects and women with eating disorders. *Journal of Psychosomatic Research, 31,* 161–169.

Wardle, J., & Beales, S. (1987). Restraint and food intake: An experimental study of eating patterns in the laboratory and in normal life. *Behaviour Research and Therapy, 25,* 179–185.

Westenhoefer, J. (1991). Dietary restraint and disinhibition: Is restraint a homogeneous construct? *Appetite, 16,* 45–55.

Westenhoefer, J., Broeckmann, P., Munch, A. K., & Pudel, V. (1994). Cognitive control of eating behaviour and the disinhibition effect. *Appetite, 23,* 27–41.

Westenhoefer, J., Stunkard, A., & Pudel, V. (1999). Validation of the flexible and rigid control dimensions of dietary restraint. *International Journal of Eating Disorders, 26,* 53–64.

Westerterp, K. R., Nicolson, N. A., Boots, J. M., Mordant, A., & Westerterp, M. S. (1988). Obesity, restrained eating and the cumulative intake curve. *Appetite, 11,* 119–128.

Westerterp-Plantenga, M. S., Kempen, K. P., & Saris, W. H. (1998). Determinants of weight maintenance in women after diet-induced weight reduction. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity, 22,* 1–6.

Westerterp-Plantenga, M. S., Wouters, L., & ten Hoor, F. (1991). Restrained eating, obesity, and cumulative food intake curves during four-course meals. *Appetite, 16,* 149–158.

Williams, A., Spencer, C. P., & Edelmann, R. J. (1987). Restraint theory, locus of control and the situational analysis of binge eating. *Personality and Individual Differences, 8,* 67–74.

Williamson, D. A., Lawson, O. J., Brooks, E. R., Wozniak, P. J., Ryan, D. H., Bray, G. A., et al. (1995). Association of body mass with dietary restraint and disinhibition. *Appetite, 25,* 31–41.

Williamson, D. A., Martin, C. K., York-Crowe, E., Anton, S. D., Redman, L. M., Han, H., et al. (2007). Measurement of dietary restraint: Validity tests of four questionnaires. *Appetite, 48,* 183–192.

Wilson, A. J., Touyz, S. W., Dunn, S. M., & Beumont, P. (1989). The Eating Behavior Rating Scale (EBRS): A measure of eating pathology in anorexia nervosa. *International Journal of Eating Disorders, 8,* 583–592.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.