MEASURING EATING DISORDER ATTITUDES AND BEHAVIORS:

A RELIABILITY GENERALIZATION STUDY

A Dissertation

by

CRYSTAL ANNE PEARSON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2006

Major Subject: Psychology

MEASURING EATING DISORDER ATTITUDES AND BEHAVIORS:

A RELIABILITY GENERALIZATION STUDY

A Dissertation

by

CRYSTAL ANNE PEARSON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Leslie C. Morey
	David H. Gleaves
Committee Members,	Winfred Arthur, Jr.
	Linda G. Castillo
Head of Department,	Leslie C. Morey

December 2006

Major Subject: Psychology

ABSTRACT

Measuring Eating Disorder Attitudes and Behaviors:

A Reliability Generalization Study. (December 2006)

Crystal Anne Pearson, B.A., Mississippi State University;

M.S., Texas A&M University

Co-Chairs of Advisory Committee: Dr. Leslie C. Morey
Dr. David H. Gleaves

I used reliability generalization procedures to determine the mean score reliability of the Eating Disorder Inventory (EDI), the Eating Attitudes Test (EAT), and the Bulimia Test (BULIT). Reliability generalization is a type of meta-analysis used to examine the mean score reliability of a measure across studies and to explore study factors that influence mean score reliability. Score reliability estimates were included in 28.67% of 293 studies using the EDI, 36.28% of 215 studies using the EAT, and 41.46% of 41 studies utilizing the BULIT. For the EDI, mean Cronbach's alphas for the subscales ranged from .52 to .89 and the mean estimate for the total score was .91. For the EAT-40 and EAT-26, mean estimates of internal consistency were .81 and .86 respectively. Mean estimates of internal consistency for the EAT-26 subscales ranged from .56 to .80. The mean estimate of internal consistency for the BULIT-R was .93. Overall, the mean reliability of scores on all three measures and their subscales/factors was acceptable except for the Asceticism subscale of the EDI and the Oral Control factor on the EAT-26, which had mean internal consistency estimates of .52 and .56 respectively. For the EDI, the majority of the subscales that measure specific eating

disorder attitudes and behaviors, such as Bulimia and Perfectionism displayed higher score reliability in clinical eating disorder samples than in nonclinical samples. This difference was not found in the Drive for Thinness and Body Dissatisfaction subscales, perhaps because these attitudes are common in both eating disorder and nonclinical samples. Score reliability information for the EAT and BULIT was primarily reported for nonclinical samples; therefore, it is difficult to characterize the effect of type of sample on these measures. There was a tendency for mean score reliability for all the measures to be higher in the adult samples than in adolescent samples and in the female samples compared to the male samples. This study highlights the importance of assessing and reporting internal consistency every time a measure is used because reliability is affected by characteristics of the participants being examined.

TABLE OF CONTENTS

Page

ABSTRACT iii

TABLE OF CONTENTS v

LIST OF TABLES vi

INTRODUCTION 1

- Measurement of Eating Disorder Attitudes and Behaviors 3
- Reliability 8
- Reliability as a Property of Test Scores 9
- Reliability Generalization 11

METHOD 15

- Test Selection 15
- Developing a Coding Form and Data Collection 15
- Identification and Use of Potential Dependent Variables 16
- Data Analyses 17

RESULTS 18

- Eating Disorder Inventory 18
- Eating Attitudes Test 25
- Bulimia Test 28
- Test-Retest Reliability Analyses 30

DISCUSSION 31

- Implications 34
- Recommendations for Future RG Studies 34

CONCLUSIONS 36

REFERENCES 37

APPENDIX 44

VITA 59

LIST OF TABLES

TABLE		Page
1	Summary of Eating Disorder Measures	5
2	List of Published Reliability Generalization Studies	13
3	Mean Internal Consistency Estimates for the Eating Disorders Inventory (EDI)	19
4	Mean Internal Consistency Estimates for the Eating Attitudes Test (EAT)	27
5	Mean Internal Consistency Estimates for the Bulimia Test-Revised (BULIT-R)	29
6	Mean Test-Retest Reliability Estimates for the EDI, EAT-26, and BULIT-R	30

INTRODUCTION

Eating disorders are characterized by disturbances in eating behavior and body image. The primary types of eating disorders are anorexia nervosa, bulimia nervosa, and binge eating disorder. Although the average prevalence rates of eating disorders are relatively small, with the rate for anorexia nervosa at .3% and the rate for bulimia nervosa around 1% when using strict diagnostic criteria (Hoek & van Hoeken, 2003), these disorders affect a large number of individuals, particularly women (J. K. Thompson, 2004). Eating disorders can have devastating consequences, both medically and psychologically, and often co-occur with other psychological disorders (Agras, 2001; J. K. Thompson, 2004). Medical complications of eating disorders are wide-ranging and may include cardiovascular problems, electrolyte abnormalities, osteoporosis, gastrointestinal problems, endocrine and metabolic abnormalities, dental problems, and infertility (Agras, 2001; Pomeroy & Mitchell, 2002). Major depressive disorder and personality disorders are comorbid psychological disorders common to all the eating disorders, whereas anxiety disorders are more commonly associated with anorexia nervosa and substance abuse disorders with bulimia nervosa and binge eating disorder (Agras, 2001).

In a factor analytic investigation, Williamson and colleagues (2002) found three latent features described the symptoms of eating disorders: binge eating, fear of fatness/compensatory behavior, and drive for thinness. Eating disorder groups were also

---

This dissertation follows the style of *Psychology of Addictive Behaviors.*

differentiated by these features, with the anorexia nervosa group scoring highly on drive for thinness and fear of fatness/compensatory behavior, the bulimia nervosa group scoring highly on binge eating and fear of fatness/compensatory behavior, and the binge eating disorder group scoring highly only on the feature of binge eating. As defined by the DSM-IV-TR, central features of anorexia nervosa include low body weight, restriction of food intake, fear of gaining weight, and distorted concerns about body shape or weight (American Psychiatric Association [APA], 2000). Other characteristics of anorexia nervosa may include amenorrhea, episodes of binge eating, and purging behavior such as self-induced vomiting, laxative misuse, or excessive exercise. Bulimia nervosa is primarily characterized by episodes of binge eating followed by inappropriate compensatory methods to prevent weight gain. As in anorexia nervosa, these compensatory methods may include self-induced vomiting, laxative abuse, excessive exercise, or fasting. Other features of bulimia nervosa include a feeling of loss of control during binge episodes and a self-evaluation that is overly influenced by body shape and weight. Binge eating disorder is characterized by repeated episodes of binge eating that are not followed by compensatory behavior (APA, 2000).

Many measures to assess eating disorders have been developed, allowing researchers and clinicians to further understand these disorders and examine the effects of treatment interventions. In order to reach these goals, researchers and clinicians must use measures that have scores that have been found to be reliable in many situations, so researchers can have confidence in the conclusions drawn based on these scores. The score reliability of eating disorder measures is important because an individual's score

on a measure may be used to guide decisions about the necessity of treatment and findings from clinical studies based upon these scores are used to guide treatment decisions. There is potential for harm to individuals if researchers or clinicians assume measures are reliable in all circumstances. The purpose of this study is to examine score reliability of several measures of eating disorder symptoms and the conditions under which scores on these measures show greater reliability.

Measurement of Eating Disorder Attitudes and Behaviors

The two most common methods for assessment of eating disorders are structured interviews and self-report measures. Examples of structured interviews for assessing attitudes and behaviors of eating disorders include the Eating Disorder Examination (EDE; Fairburn & Cooper, 1993) and the Interview for the Diagnosis of Eating Disorders (IDED; Kutlesic, Williamson, Gleaves, Barbin, & Murphy-Eberenz, 1998). The EDE uses symptom ratings to assess both anorexia nervosa and bulimia nervosa and contains four subscales: restraint, eating concern, shape concern, and weight concern. The IDED assesses anorexia nervosa, bulimia nervosa, and binge eating disorder. Similar to the EDE, interviewers using the IDED make ratings on 21 symptoms, and these ratings are used for diagnosis. The EDE and a self-report measure based on the EDE produced similar results with respect to the evaluation of clear-cut behavioral features of eating disorders such as vomiting and laxative abuse, but compared to the EDE ratings, scores on the self-report measure indicated greater reporting of problems with respect to the more complex aspects of eating disorders, such as binge eating and body image (Fairburn & Beglin, 1994). Garner (1995) argued that self-report measures are less accurate than

interviews regarding a behavior such as binge eating that may be subjectively defined by the participant; however, self-report measures do have some advantages when compared to interviews. They are relatively brief, easy to administer, and economical compared to an interview. They are also objectively scored which eliminates possible bias from the interviewer (Garner, 1995). Self-report measures are not "personally intrusive," which may make it easier to admit to behaviors such as binge-eating and self-induced vomiting that might be considered embarrassing (Fairburn & Beglin, 1994, p. 364).

Three commonly used self-report measures of bulimic attitudes and behaviors are the Eating Disorder Inventory (EDI; Garner, Olmstead, & Polivy, 1983), the Eating Attitudes Test (EAT; Garner & Garfinkel, 1979), and the Bulimia Test (BULIT; Smith & Thelen, 1984). See Table 1 for a summary of these measures. As self-report measures are widely used in both clinical and research settings, it is important that the reliability of the scores obtained from these measures be well-established, and that professionals using these tests understand the conditions affecting the reliability of scores on these measures.

The Eating Disorders Inventory (EDI; Garner et al., 1983) is a self-report measure of both anorexia nervosa and bulimia nervosa. The original EDI contains 64 items that assess both behavioral and psychological traits of anorexia nervosa and bulimia nervosa. It consists of the following eight subscales: Drive for Thinness, Bulimia, Body Dissatisfaction, Ineffectiveness, Perfectionism, Interpersonal Distrust, Interoceptive Awareness, and Maturity Fears. The first three subscales assess attitudes and behaviors related to eating, weight, and body shape, and the last five subscales

Table 1

Summary of Eating Disorder Measures

Measure	Versions	# of Items	# of Subscales/Factors	Disorder(s) Assessed
Eating Disorder Inventory (EDI)	EDI: (Garner, Olmstead, & Polivy, 1983)	64	8	Anorexia nervosa and bulimia
	EDI-2: (Garner, 1991)	91	11	nervosa
Eating Attitudes Test (EAT)	EAT: (Garner & Garfinkel, 1979)	40	--	Anorexia nervosa and bulimia
	EAT-26: (Garner, Olmsted, Bohr, & Garfinkel, 1982)	26	3	nervosa
Bulimia Test (BULIT)	BULIT: (Smith & Thelen, 1984)	32	--	Bulimia nervosa
	BULIT-R: (Thelen, Farmer, Wonderlich, & Smith, 1991)	28	--	

measure general psychological traits believed to be related to eating disorders (Garner, 1995). According to Garner and colleagues (1983), the Drive for Thinness subscale evaluates preoccupation with weight and an extreme concern with dieting. The Bulimia subscale assesses for the presence of episodes of binge eating followed by a desire to purge by self-induced vomiting. The Body Dissatisfaction subscale examines whether an individual is dissatisfied by body shape and believes certain body parts to be too large or too fat. The Ineffectiveness subscale evaluates general feelings of inadequacy and worthlessness, while the Perfectionism subscale assesses for extreme expectations of

superior performance and perfectionistic tendencies. Feelings of estrangement and reluctance to form close relationships are measured by the Interpersonal Distrust subscale, and the Interoceptive Awareness subscale evaluates difficulty identifying emotions and feelings of hunger. The Maturity Fears subscale examines a desire to escape the demands of being an adult by withdrawing to childhood years (Garner et al., 1983).

The revised version of this measure, the EDI-2 (Garner, 1991), contains 91 items measuring 11 subscales–eight subscales from the original EDI and three provisional subscales measuring Asceticism, Impulse Regulation, and Social Insecurity (Williamson, Anderson, & Gleaves, 1996). The Asceticism subscale examines a need for self-discipline, the Impulse Regulation subscale assesses impulsiveness, and the Social Insecurity subscale evaluates the belief that social relationships are insecure and unrewarding (Thurfjell et al., 2004). In a clinical sample, Eberenz and Gleaves (1994) reported generally high internal consistency (Cronbach's alphas ranging from .80 to .91) for scores on the eight original scales, but alphas for the three new scales were all found to be less than .80. Eberenz and Gleaves (1994) found support for the factor structure of the original eight scales in a clinical sample, but the researchers did not find support for the proposed factor structure of the additional subscales. Joiner and Heatherton (1998) also found support for the factor structure of the eight original EDI scales in a non-clinical sample.

The EDI was principally designed to assess anorexia, with bulimic attitudes and behaviors considered to be primarily a subtype of anorexia, although it was recognized

that bulimia could occur without a history of anorexia (Garner et al., 1983). Based on this conceptualization, Garner et al. (1983) found that the EDI distinguished anorexics from controls and it also distinguished what the researchers referred to as purging and nonpurging anorexics. In a later study, Gross, Rose, Leitenberg, and Willmuth (1986) reported that the EDI discriminated between control participants and a clinical group of bulimia nervosa patients.

Garner and Garfinkel (1979) developed the Eating Attitudes Test (EAT) to evaluate thoughts and behaviors related to anorexia nervosa. There are two versions of this measure. The original version consisted of 40 items scored on a 6-point Likert scale ranging from "always" to "never" (Williamson et al., 1996). Garner and colleagues (1982) revised the EAT into a 26-item version, the EAT-26, based on a factor analysis of the original form, and Scheinberg et al. (1993) reported a .98 correlation between the short and long forms. Although the EAT was originally developed as an assessment of symptoms related to anorexia nervosa, it has also been shown to discriminate individuals with bulimia nervosa from control participants (Gross et al., 1986). The EAT has been found to distinguish eating disordered patients and controls and also to differentiate binge eating patients from anorexic and bulimic patients; however, it has not been shown to distinguish anorexic patients from bulimic patients (Williamson et al., 1996).

The EAT-26 consists of 3 factors: Dieting, Bulimia and Food Preoccupation, and Oral Control (Garner, Olmstead, Bohr, & Garfinkel, 1982). The Dieting factor evaluates the "avoidance of fattening foods and a preoccupation with being thinner" (Garner et al., 1982, p. 873). The Bulimia and Food Preoccupation factor assesses binge eating and

purging, as well as thoughts about food. The Oral Control factor examines an individual's ability to control food intake and perceived pressure from other people to gain weight (Garner et al., 1982). According to Raciti and Norcross (1987), the EDI evaluates more of the cognitive and behavioral characteristics of eating disorders compared to the EAT which primarily assesses behavioral features of the disorders. The strongest correlations between the two were found with the overall EAT score and scores on the three eating disorder symptom scales of the EDI (drive for thinness, body dissatisfaction, and bulimia) (Gross et al., 1986).

Smith and Thelen (1984) developed the Bulimia Test (BULIT) because they believed that the lack of an appropriate scale to assess symptoms of bulimia was impairing both research and clinical work in this area. The BULIT is a 32-item, multiple-choice inventory constructed by comparing the responses of a clinical sample of bulimic subjects with normal female college students on 75 questions based on the DSM-III criteria for bulimia. In 1991, the BULIT was revised to be consistent with criteria for bulimia nervosa in the DSM-III-R (Thelen, Farmer, Wonderlich, & Smith, 1991). The revised measure contained 28 scored items and was highly correlated with the original version (Williamson et al., 1996). The BULIT-R has also been found to measure the symptoms of bulimia nervosa as defined by DSM-IV criteria (Thelen, Mintz, & Vander Wal, 1996).

Reliability

Reliability refers to the stability of an individual's test scores over repeated administrations of a test or alternate forms of a test (Yin & Fan, 2000). In classical test

theory, an individual's observed score is made up of his or her true score and error (Eason, 1991; Henson & B. Thompson, 2002). Test score reliability can be estimated in many ways, but the most widely used method is a measure of the internal consistency of a measure, Cronbach's (1951) coefficient alpha (Hogan, Benjamin, & Brezinski, 2003). Other ways to measure reliability include examining temporal stability estimates, or test-retest reliability, and equivalency estimates, such as the correlation between alternate forms of a measure (American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME], 1999). Each of these estimates considers only one source of error, and classical test theory does not examine interactions among error sources (Eason, 1991). One advantage of an internal consistency estimate is that the measure only needs to be given once (B. Thompson, 2003). This advantage also means that researchers should nearly always be able to calculate and report an estimate of internal consistency.

Reliability as a Property of Test Scores

When reporting their research, authors of behavioral research studies frequently refer to the "reliability of a measure." According to B. Thompson (1994), use of this shorthand phrase contributes to the misunderstanding by many researchers and students about reliability, as it is the scores on a measure that are found to be reliable. Wilkinson and the APA Task Force on Statistical Inference (1999) emphasized this principle by stating that "a test is not reliable or unreliable" instead "reliability is a property of the scores on a test for a particular population of examinees" (p. 596). This distinction is an important one, as reliability coefficients may vary depending on characteristics of the

sample. Wilkinson and colleagues (1999) proposed guidelines for reporting statistical methods in research articles and emphasized that the psychometric properties of a measure should always be reported, stressing that "authors should provide reliability coefficients of the scores for the data being analyzed even when the focus of their research is not psychometric" (p. 596). This statement emphasizes the importance of reporting reliability data whenever a test is used regardless of the primary intent of the article. Unfortunately, many researchers and/or journal editors seem to be unaware of these guidelines and do not report reliability coefficients, perhaps because they believe that a measure was proven reliable during test construction or because they believe that score reliability does not influence the results of their study.

There are several reasons why it is important to examine and report reliability of test scores every time a measure is used. First, if score reliability is poor, the ability to measure the intended construct may be compromised leading to a potential problem with validity of the data (AERA/APA/NCMA, 1999; B. Thompson, 2003). Reliability of test scores is a necessary condition to establish validity, as "unreliable scores measure nothing" (B. Thompson, 2003, p. 6). Additionally, poor score reliability may also hinder the ability to find statistically, clinically, or practically significant effects (B. Thompson, 2003). When interpreting effect sizes, score reliability is an important factor to consider because measurement error impacts effect size (Baugh, 2002; B. Thompson, 2001; Vacha-Haase, 1998), as a larger standard error contributes to a less precise effect size value (Lipsey & Wilson, 2001). Measurement errors cause observed effects to fluctuate across studies and may lead to underestimation of true effects (Baugh, 2002), which has

led to recommendations for correcting effect size estimates for unreliable scores (Hunter & Schmidt, 1994). Finally, total score variance affects reliability of the data set, and total score variance is impacted by characteristics of the participants (B. Thompson, 1994; Yin & Fan, 2000). Because score variability is a property of the data, reliability estimates will not remain constant across studies and should be therefore be evaluated and reported as part of the process of describing the data.

Given the importance of test score reliability to scientific studies, it is surprising that the editorial policies of journals do not require this information to be reported. Although some journals, such as *Educational and Psychological Measurement*, now require the reporting of reliability information (B. Thompson, 2003), most journals still do not require the inclusion of this information and most authors do not report reliability estimates for their data (Meier & Davis, 1990; Vacha-Haase, Henson, & Caruso, 2002).

Reliability Generalization

One method of describing measurement error in a test's scores across studies is reliability generalization (RG). RG, a type of meta-analysis, characterizes the typical (i.e., mean) reliability of scores across studies, the amount of variability in reliability coefficients, and the sources of variability in reliability coefficients (Vacha-Haase, 1998). RG is consistent with previous work on validity generalization (Hunter & Schmidt, 1990), in which researchers conducted analyses to determine if the validity of scores on a test was generalizable to different samples (Vacha-Haase, 1998; Viswesvaran & Ones, 2000). As with other types of meta-analysis, RG allows researchers to understand a large body of literature which may be producing inconsistent findings, in

this case helping to understand differences in score reliability across multiple studies (Yin & Fan, 2000).

By examining factors that influence the reliability of scores across studies, we can better understand factors that contribute to variability in score quality. The factors that influence score reliability are typically dependent on characteristics of the participants from which the data are gathered. For example, other factors being equal, a heterogeneous set of participants will produce higher score reliability than a more homogenous groups of participants (Yin & Fan, 2000). Other participant characteristics to consider include the type of sample, as well as age, gender, and ethnicity of participants. Other study factors potentially impacting test score reliability are sample size, type of reliability, culture, test format, test length, and test language. By explicating the conditions under which a test's scores display higher or lower reliability, researchers will be able to tailor future studies to conditions that will maximize score reliability. By maximizing score reliability, researchers will have additional control over one factor that influences effect sizes. For example, if a researcher is conducting a study in a clinical population and the researcher knows that scale A of a construct has significantly higher score reliability then scale B in clinical samples, he or she could select scale A for use in the study.

Although RG studies are only beginning to be widely conducted, this method has been employed to examine the reliability of scores on measures such as the Minnesota Multiphasic Personality Inventory, the Beck Depression Inventory (BDI; Beck & Steer, 1984), and the NEO-Five Factor Inventory. See Table 2 for a listing of published RG

Table 2

List of Published Reliability Generalization Studies

Measure	**Study**
Adult attachment style measures	Reese, Kieffer, & Briggs (2002)
Alcohol Use Disorders Identification Test	Shields & Caruso (2003)
Beck Depression Inventory	Yin & Fan (2000)
Bem Sex Role Inventory	Vacha-Haase (1998)
"Big Five Factors" of Personality	Viswesvaran & Ones (2000)
CAGE Questionnaire	Shields & Caruso (2004)
Career Decision-Making Self-Efficacy Scale	Nilsson, Schmidt, & Meek (20020
Coopersmith Self-Esteem Inventory	Lane, White, & Henson (2002)
Differential Emotions Scale	Youngstrom & Green (2003)
Eysenck Personality Questionnaire	Caruso, Witkiewitz, Belcourt-Dittloff, & Gottlieb (2001)
Geriatric Depression Scale	Kieffer, Reese, & Kieffer (2002)
Junior Eysenck Personality Questionnaire	Caruso & Edwards (2001)
Kolb's Learning Style Inventory	Henson & Hwang (2002)
LibQUAL+-super ™	Thompson & Cook (2002)
Life Satisfaction Index	Wallace & Wheeler (2002)
Marlowe-Crowne Social Desirability Scale	Beretvas, Meyers, & Leite (2002)
Mathematics Anxiety Rating Scale	Capraro, Capraro, & Henson (2001)
MMPI (Clinical Scales)	Vacha-Haase, Kogan, Tani, & Woodall (2001)
MMPI/MMPI-2 (Validity Scales)	Vacha-Haase, Tani, Kogan, Woodall, & Thompson (2001)
Myers-Briggs Type Indicator	Capraro & Capraro (2002)
NEO personality scales	Caruso (2000)
Spielberger State-Trait Anxiety Inventory	Barnes, Harp, & Jung (2002)
Teacher Efficacy Scale	Henson, Kogan, & Vacha-Haase (2001)

studies and the measures examined in these studies. As mentioned above, reliability estimates still appear to be rarely reported, as indicated by Yin and Fan's (2000) finding of a 7.5% reporting rate for the BDI and Caruso's (2000) finding of a 15.2% reporting

rate for the NEO. Henson and B. Thompson (2002) suggest that reporting rates are unlikely to exceed 40% for any test. Low rates of reliability reporting are perhaps the greatest hindrance to conducting an RG study and may explain why this method has not been more widely employed.

The goals of this study were to use reliability generalization procedures to report the mean score reliability for the EDI, the EAT, and the BULIT and to explore how score reliability of these eating disorder measure varies across studies and what study characteristics account for this variation. Presently, the author did not identify any published RG studies on the score reliability of any measure of eating disorders or related constructs. This RG analysis of selected eating disorder measures advances the field by providing information on factors that influence reliability on this type of measure, as well as providing information about the typical score reliability of these measures. This study advances the field of eating disorder research by providing information on measurement issues that influence substantive findings. Additionally, RG is also a valuable method of educating other researchers about reliability issues and emphasizing that reliability is “not an immutable unchanging property of tests” (Henson & B. Thompson, 2002, p. 124).

METHOD

I followed the framework proposed by Henson and B. Thompson (2002) for designing an RG study. This framework consists of the following five steps: selecting the measures to be analyzed, developing a coding sheet, collecting data, identifying potential dependent variables, and conducting analyses.

Test Selection

I analyzed studies that utilize forms of the EDI, the EAT, and/or the BULIT. These measures were selected because they are probably the most commonly used measures of eating disorder symptomatology and many studies have been published utilizing these measures. Additionally, I reviewed studies of the EDE and the IDED-IV to determine if enough information was available on these interviews to conduct RG analyses. Because little score reliability data were available, these measures were not included for analysis. For example, only five articles were found in a PsycINFO search for the IDED and two of these articles were not in English.

Developing a Coding Form and Data Collection

Relevant reports were gathered through database searches of PsycINFO using the terms Eating Disorder Inventory, Eating Attitudes Test, and Bulimia Test. Searches resulted in 873 references for the EDI, 601 for the EAT, and 131 for the BULIT. The inclusion criterion was published empirical journal articles, as the goal of the current project was to evaluate the score reliability information reported in the literature. Article sources that were excluded from further study included books/book chapters, theoretical

articles, review articles, case studies, dissertations, meta-analyses, and articles not published in English.

As a result of the inclusion criterion, 293 studies of the EDI, 214 studies of the EAT, and 41 studies of the BULIT were reviewed and coded. These studies were coded based on whether reliability information for the sample was reported and what type of reliability information was provided (i.e., internal consistency or stability). Additional study factors were also coded, including type of reliability coefficient reported, type of sample (clinical–eating disorder, clinical–general psychiatric, nonclinical, or mixed sample), type of study (treatment or other), age of participants, gender of participants, test language, test form, test length, and sample size. A single coder, the author, was used to code all studies. The decision to use a single coder was made because, unlike in more traditional meta-analyses, this study did not require any calculations to be made.

Identification and Use of Potential Dependent Variables

I conducted separate analyses using internal consistency and test-retest coefficients as dependent variables. When using reliability estimates, some researchers combine Cronbach's alpha estimates with test-retest reliability estimates as a single dependent variable, but Dimitrov (2002) cautions against combining these estimates as they are not equivalent and combining them together could lead to "mixing apples and oranges" (Dimitrov, 2002, p. 792). After coding all the studies, I determined which study features were used in the analyses as independent variables based on whether enough data were available using that feature.

Data Analyses

SAS 8.2 software was used for all analyses, and I followed the guidelines suggested by Henson and B. Thompson (2002) for conducting a reliability generalization study and Arthur, Bennett, and Huffcutt (2001) for conducting a meta-analysis. Overall mean reliability coefficients weighted by sample size were calculated for each measure, and sample-weighted mean reliability estimates broken down by predictor variables were also calculated. Sample size is one source of sampling error, with larger sample sizes providing more stable estimates of the population parameter because they are less susceptible to sampling error than smaller samples. Therefore, sample weighted means were used to reduce the effects of sampling error from smaller samples. If data were available, mean reliability coefficients are reported for the subscales of the measures. Additionally, the 95% confidence interval and percent of variance accounted for by sampling error were also calculated.

RESULTS

Eating Disorder Inventory

In 155 (52.90%) out of 293 studies including the EDI, the researchers did not provide score reliability information. In 54 (18.43%) studies, the researchers cited reliability estimates of scores from previously published studies or stated that the measure had been found to be reliable. The researchers reported reliability information for either the total scale score or one or more subscale scores for their sample in 84 (28.67%) studies; however, 10 of these studies were excluded from the analyses because the authors only reported a range of reliability coefficients for the subscale scores and 9 studies were excluded for using a different factor structure. Five studies included only test-retest reliability and were analyzed separately. In some studies, researchers reported reliability information for more than one group of participants (e.g., control group and clinical group), resulting in more reliability coefficients for that scale than there were studies reporting score reliability information for that scale. See the Appendix for the references that reported score reliability information for their study sample.

Table 3 presents the mean estimates of internal consistency for the EDI and its subscales and the means broken down by gender and age of participants, as well as type of sample and test language. I was not able to analyze all of the coded study factors due to low variability or insufficient reporting of the characteristic. Only one of the articles reporting score reliability information was a treatment study, and the majority of authors did not provide sufficient information regarding participant ethnicity to allow for further analysis.

Table 3

Mean Internal Consistency Estimates for the Eating Disorders Inventory (EDI)

Study Characteristics		*Reliability*				
		# of Data Points K	**Total Sample Size N**	**Sample-weighted Mean Alpha (SD)**	**95% CI**	**PVA_{SE}**
EDI Total		**6**	**3269**	**.91 (.06)**	**.86 - .95**	**1.65%**
Gender:	Female	6	3269	.91 (.06)	.86 - .95	1.65%
Age:	Adult	4	995	.85 (.08)	.77 - .93	4.35%
	Adolescent	2	1137	.93 (.003)	.93 - .93	100.00%
Type of Sample:	Nonclinical	4	2989	.90 (.06)	.84 - .96	1.24%
	Clinical: Eating Disorder	2	280	.95 (.009)	.93 - .96	99.32%
Test Language:	English	4	995	.85 (.08)	.77 - .93	4.35%
	Non-English	2	2274	.93 (.002)	.93 - .93	100.00%
EDI Body Dissatisfaction Subscale		**55**	**22120**	**.89 (.04)**	**.88 - .90**	**6.51%**
Gender:	Female	44	17240	.90 (.03)	.89 - .91	9.82%
	Male	5	2443	.81 (.03)	.79 - .84	26.56%
	Mixed gender	6	2437	.90 (.03)	.88 - .92	11.02%
Age:	Adult	35	8311	.91 (.03)	.90 - .92	17.16%
	Adolescent	20	13809	.88 (.04)	.86 - .90	3.92%
Type of Sample:	Nonclinical	43	19844	.89 (.04)	.87 - .90	5.63%
	Clinical: Eating Disorder	11	2156	.90 (.03)	.89 - .92	19.05%
Test Language:	English	42	13557	.89 (.03)	.88 - .90	10.70%
	Non-English	13	8563	.88 (.05)	.85 - .91	3.17%
EDI Drive for Thinness Subscale		**49**	**22335**	**.85 (.05)**	**.83 - .86**	**6.24%**
Gender:	Female	37	15831	.84 (.05)	.83 - .86	6.56%
	Male	4	2277	.79 (.03)	.77 - .82	37.17%
	Mixed gender	8	4227	.89 (.02)	.87 - .90	27.54%
Age:	Adult	32	7784	.86 (.04)	.85 - .88	15.07%
	Adolescent	17	14551	.84 (.06)	.81 - .86	3.29%
Type of Sample:	Nonclinical	36	19788	.85 (.05)	.83 - .87	4.78%
	Clinical: Eating Disorder	12	2427	.83 (.04)	.80 - .85	37.93%
Test Language:	English	37	13943	.86 (.05)	.84 - .87	8.58%
	Non-English	12	8392	.83 (.06)	.80 - .86	4.17%
EDI Bulimia Subscale		**47**	**21875**	**.75 (.07)**	**.73 - .77**	**9.47%**
Gender:	Female	37	15905	.77 (.07)	.74 - .79	9.18%
	Male	3	2081	.67 (. 02)	.65 - .69	100.00%
	Mixed gender	7	3889	.74 (.04)	.71 - .77	20.89%
Age:	Adult	31	7437	.81 (.07)	.79 - .83	11.55%
	Adolescent	16	14438	.72 (.04)	.70 - .74	14.17%
Type of Sample:	Nonclinical	35	19861	.74 (.06)	.72 - .76	9.98%
	Clinical: Eating Disorder	11	1894	.84 (.07)	.79 - .88	10.76%
Test Language:	English	35	13483	.76 (.07)	.74- .78	10.71%
	Non-English	12	8392	.74 (.06)	.70 - .77	7.57%

Table 3 Continued

Study Characteristics		*Reliability*				
		# of Data Points K	**Total Sample Size N**	**Sample-weighted Mean Alpha (SD)**	**95% CI**	**PVA_{SE}**
EDI Perfectionism Subscale		**34**	**15888**	**.68 (.08)**	**.65 - .70**	**8.79%**
Gender:	Female	28	13487	.68 (.09)	.65 - .72	7.72%
	Male	4	2191	.63 (.05)	.58 - .68	26.91%
Age:	Adult	23	4907	.75 (06)	.73 - .78	24.42%
	Adolescent	11	10981	.64 (.07)	.60 - .68	7.19%
Type of Sample:	Nonclinical	22	13716	.66 (.08)	.63 - .70	7.93%
	Clinical: Eating Disorder	11	2052	.77 (.06)	.73 - .80	25.04%
Test Language:	English	22	7496	.68 (.12)	.63 - .73	6.47%
	Non-English	12	8392	.67 (.04)	.65 - .70	24.19%
EDI Interoceptive Awareness Subscale		**31**	**15272**	**.75 (.06)**	**.73 - .77**	**10.44%**
Gender:	Female	26	12981	.76 (.06)	.74 - .78	10.70%
	Male	3	2081	.69 (.03)	.65 - .72	36.68%
Age:	Adult	20	4606	.80 (.07)	.77 - .83	13.17%
	Adolescent	11	10666	.73 (.05)	.70 - .76	10.88%
Type of Sample:	Nonclinical	19	13314	.74 (.06)	.71 - .77	8.88%
	Clinical: Eating Disorder	10	1781	.81 (.04)	.79 - .84	51.53%
Test Language:	English	19	6880	.78 (.06)	.75 - .80	10.93%
	Non-English	12	8392	.73 (.05)	.70 - .76	12.81%
EDI Ineffectiveness Subscale		**29**	**14683**	**.80 (.06)**	**.77 - .82**	**6.54%**
Gender:	Female	24	12392	.80 (.07)	.78 - .83	5.53%
	Male	3	2081	.76 (.01)	.75 - .77	100.00%
Age:	Adult	19	4074	.85 (.05)	.82 - .87	12.94%
	Adolescent	10	10609	.78 (.06)	.74 - .81	4.66%
Type of Sample:	Nonclinical	18	12782	.79 (.06)	.76 - .82	5.71%
	Clinical: Eating Disorder	10	1781	.87 (.03)	.85 - .89	34.16%
Test Language:	English	17	6291	.78 (.08)	.74 - .82	5.87%
	Non-English	12	8392	.81 (.04)	.79 - .83	11.30%
EDI Maturity Fears Subscale		**31**	**14963**	**.69 (.09)**	**.66 - .72**	**7.61%**
Gender:	Female	25	12506	.69 (.09)	.65 - .72	7.25%
	Male	4	2247	.69 (.08)	.62 - .77	8.44%
Age:	Adult	20	4240	.77 (.08)	.73 - .80	11.39%
	Adolescent	11	10723	.66 (.07)	.62 - .70	7.49%
Type of Sample:	Nonclinical	20	13062	.67 (.08)	.64 - .71	8.02%
	Clinical: Eating Disorder	10	1781	.82 (.05)	.79 - .85	24.25%
Test Language:	English	19	6571	.66 (.11)	.61 - .71	7.21%
	Non-English	12	8392	.71 (.05)	.69 - .74	14.88%

Table 3 Continued

Study Characteristics		*Reliability*				
		# of Data Points K	**Total Sample Size N**	**Sample-weighted Mean Alpha (SD)**	**95% CI**	**PVA_{SE}**
EDI Interpersonal Distrust Subscale		**29**	**14683**	**.71 (.08)**	**.68 - .74**	**6.90%**
Gender:	Female	24	12392	.72 (.09)	.68 - .75	5.90%
	Male	3	2081	.68 (.04)	.63 - .72	32.00%
Age:	Adult	19	4074	.78 (06)	.75 - .81	21.00%
	Adolescent	10	10609	.69 (.08)	.64 - .73	4.48%
Type of Sample:	Nonclinical	18	12782	.70 (.08)	.66 - .73	5.73%
	Clinical: Eating Disorder	10	1781	.80 (.03)	.78 - .82	74.12%
Test Language:	English	17	6291	.69 (.11)	.63 - .74	5.90%
	Non-English	12	8392	.73 (.04)	.71 - .75	18.59%
EDI Asceticism Subscale*		**12**	**5937**	**.52 (.10)**	**.46 - .58**	**10.12%**
Gender:	Female	9	4029	.58 (.07)	.53 - .62	19.29%
Age:	Adult	9	1965	.61 (.07)	.57 - .66	32.71%
	Adolescent	3	3972	.47 (.08)	.38 - .57	6.78%
Type of Sample:	Nonclinical	5	4606	.49 (.09)	.41 - .57	7.22%
	Clinical: Eating Disorder	6	1211	.62 (.07)	.57 - .68	41.89%
Test Language:	English	3	596	.60 (.05)	.55 - .66	90.15%
	Non-English	9	5341	.51 (.10)	.44 - .58	8.58%
EDI Impulse Regulation Subscale*		**13**	**6241**	**.76 (.05)**	**.74 - .79**	**15.85%**
Gender:	Female	10	4333	.75 (.05)	.72 - .79	15.13%
Age:	Adult	10	2269	.74 (.07)	.70 - .79	17.46%
	Adolescent	3	3972	.78 (.01)	.77 - .79	100.00%
Type of Sample:	Nonclinical	6	4910	.78 (.03)	.75 - .80	17.10%
	Clinical: Eating Disorder	6	1211	.72 (.07)	.67 - .78	24.21%
Test Language:	English	4	900	.78 (.03)	.75 - .81	75.32%
	Non-English	9	5341	.76 (.05)	.73 - .79	12.10%
EDI Social Insecurity Subscale*		**13**	**6327**	**.72 (.04)**	**.70 - .75**	**28.07%**
Gender:	Female	10	4419	.72 (.05)	.69 - .75	24.36%
Age:	Adult	10	2355	.75 (.04)	.73 - .78	44.83%
	Adolescent	3	3972	.71 (.03)	.67 - .74	23.59%
Type of Sample:	Nonclinical	6	4910	.72 (.04)	.69 - .75	17.81%
	Clinical: Eating Disorder	6	1297	.74 (.04)	.71 - .77	68.49%
Test Language:	English	4	986	.77 (.03)	.74 - .80	73.99%
	Non-English	9	5341	.71 (.04)	.69 - .74	30.03%

Note. *These subscales were added to the Eating Disorder Inventory-2. CI = Confidence Interval. PVA_{SE} = Percent of variance accounted for by sampling error.

Mean estimates of internal consistency for scores on the subscales ranged from .52 to .89 and for scores on the EDI, the mean estimate was .9. No studies reported estimates of internal consistency for the total score on the EDI-2. All analyses that utilize only a few data points will be less stable statistically than those analyses that include a higher number of data points. Therefore, some of the comparisons noted below should be interpreted with caution.

The lowest estimates of internal consistency were for the Perfectionism, Maturity Fears, and Asceticism subscales; however, only the Asceticism subscale had questionable score reliability with an estimate of .52. The Bulimia, Perfectionism, and Maturity Fears subscales had higher score reliability in clinical eating disorder samples than in nonclinical samples. The Drive for Thinness, Impulse Regulation, and Body Dissatisfaction subscales did not display this difference. Mean estimates of score reliability also tended to be higher in the adult samples compared to the adolescent samples, as well as the female samples compared to the male samples.

For the total EDI, reliability was higher in the adolescent samples compared to the adult samples, with the mean for the adult samples not falling within the 95% confidence interval of the adolescent samples. Mean estimates of reliability were also higher in the clinical eating disorder samples than the nonclinical samples, with the mean of the nonclinical samples falling below the lower bound of the 95% confidence interval for the clinical samples.

For the Body Dissatisfaction subscale, the mean estimate of internal consistency was .89. Mean score reliability was higher in the female and mixed gender samples than

in the male samples. The means of the female and mixed gender groups were not included in the confidence interval for the male samples. For this subscale, the adult samples had greater reliability than the adolescent samples. There was no difference between the clinical and nonclinical samples or between the test language samples.

The mean estimate of internal consistency for the EDI Drive for Thinness subscale was .85. Reliability was highest in the mixed gender samples followed by the female and male samples. For the other study characteristics, the adult and clinical samples displayed score reliability similar to their comparison samples, and the English samples displayed greater reliability in their scores than the non-English test language samples.

The score reliability estimate for the Bulimia subscale was .75. The female and mixed gender samples displayed greater reliability than the male samples with the confidence interval for the male samples not including the means of the other two categories. The adult estimate was also greater than the score reliability estimate for the adolescent samples and the clinical eating disorder samples were greater than the nonclinical samples.

For the Perfectionism subscale, the mean estimate of internal consistency was .68. The mean estimate for the female samples was higher than the estimate for the male samples, and the estimate for the adult samples was greater than the adolescent samples. The clinical samples also displayed greater reliability than the nonclinical samples, while the test language groups displayed nearly identical score reliability.

The mean estimate of internal consistency for the Interoceptive Awareness subscale was .75. The female and adult samples displayed greater reliability than the male and adolescent samples, respectively. The clinical eating disorder samples had higher reliability than the nonclinical samples, while the English samples had a greater mean estimate than the non-English test language samples. The Ineffectiveness subscale had a mean estimate of internal consistency of .80. The female, adult, and clinical eating disorder samples displayed greater reliability than their respective comparison samples for the Ineffectiveness subscale.

The Maturity Fear subscale had a mean estimate of .69. The female and male samples displayed similar levels of reliability. The adult samples and the clinical eating disorder samples displayed greater reliability than the adolescent and nonclinical samples, respectively. The English samples also displayed a higher mean estimate of reliability than the non-English test language sample. For the Interpersonal Distrust subscale, the mean estimate of internal consistency was .71. The adult, clinical eating disorder, and English samples had higher reliability than the male, adolescent, nonclinical, and non-English samples.

The remaining subscales were added to the revised EDI-2. As previously noted, the Asceticism subscale had the lowest mean estimate of internal consistency at .52. The adult samples displayed greater reliability than the adolescent samples and the clinical eating disorder samples had greater reliability than the nonclinical samples. For the Impulse Regulation subscale, the mean estimate was .76. For this subscale, the adolescent samples displayed higher score reliability than the adult samples, as did the

nonclinical and English samples with their respective comparison samples. Finally, for the Social Insecurity subscale, the mean estimate of internal consistency was .72. The adult reliability was higher than the adolescent samples, but the clinical eating disorder samples were not significantly higher than the nonclinical samples. The mean estimate of score reliability for the English test language samples was also was higher than the non-English samples.

The percent of variance explained by sampling error varied widely among the EDI and its subscales, ranging from 1% to 100%. Generally, analyses conducted with a smaller number of data points frequently had a greater percentage of variance accounted for by sampling error. A smaller percentage of variance accounted for by sampling error suggests a greater percentage of variance is accounted for by true score variance across the observed studies.

Eating Attitudes Test

The researchers did not provide reliability information in 93 (43.26%) of 215 studies utilizing the EAT. In 44 (20.47%) studies, the researchers made some reference to score reliability from previously published studies. The researchers reported reliability information for either the total EAT score or one or more factor scores for their sample in 78 (36.28%) studies. Seven of these studies were excluded from further analysis because the authors modified the measure or used different factors based on their own factor analysis of the EAT. As these researchers did not modify the EAT in a consistent manner, it was not possible to use these studies in further analyses.

The sample-weighted mean estimates of internal consistency were .81 for the EAT-40 and .86 for the EAT-26. The mean estimates of internal consistency for scores on the EAT-26 factors were .80 for the Dieting factor, .67 for the Bulimia and Food Preoccupation factor, and .56 for the Oral Control factor. Table 4 presents the sample-weighted mean estimates of internal consistency for the EAT, as well as the means broken down by gender, age, type of sample, and test language.

For the EAT-40, the female samples had higher reliability than the male and mixed gender samples, and the mixed gender group displaying higher reliability than the male group. Again, with a small number of data points for analyses, it is important to interpret these comparisons with caution. The adult and clinical eating disorder samples displayed greater reliability than their respective comparison samples.

For the EAT-26, the reliability was similar among the gender and age categories. The clinical eating disorder samples displayed greater reliability than the nonclinical samples, with the mixed clinical and nonclinical samples also displaying greater reliability than the nonclinical samples. The English samples also had a higher mean estimate of internal consistency than the non-English samples.

The female and male samples also displayed similar score reliability on the EAT-26 Dieting Factor. The adult samples had greater reliability than the adolescent samples, and the English test language samples had a higher mean estimate of reliability than the non-English samples. For the Bulimia and Food Preoccupation subscale, the mean estimate of reliability for the adult samples had higher reliability than the adolescent samples. For the Oral Control factor, the score reliability for the male samples was

Table 4

Mean Internal Consistency Estimates for the Eating Attitudes Test (EAT)

Study Characteristics		*Reliability*				
		# of Data Points K	**Total Sample Size N**	**Sample-weighted Mean Alpha (SD)**	**95% CI**	**PVA_{SE}**
EAT-40		**15**	**3925**	**.81 (.09)**	**.77 - .86**	**5.89%**
Gender:	Female	10	1950	.86 (.04)	.84 - .89	19.54%
	Male	2	492	.68 (.06)	.59 - .77	18.90%
	Mixed gender	3	1483	.79 (.08)	.70 - .88	4.55%
Age:	Adult	11	1855	.87 (.05)	.84 - .90	13.32%
	Adolescent	4	2070	.76 (.08)	.68 - .84	5.07%
Type of Sample:	Nonclinical	12	3538	.80 (.08)	.75 - .85	6.45%
	Clinical: Eating Disorder	2	193	.90 (.05)	.83 - .97	16.45%
Test Language:	English	11	2833	.82 (.08)	.78 - .87	6.47%
	Non-English	4	1092	.78 (.10)	.69 - .88	5.72%
EAT-26		**54**	**11963**	**.86 (.05)**	**.84 - .87**	**11.50%**
Gender:	Female	42	9566	.85 (.05)	.83 - .87	11.06%
	Male	4	488	.85 (.03)	.82 - .88	72.61%
	Mixed gender	8	1909	.87 (.05)	.84 - .90	10.28%
Age:	Adult	42	9049	.86 (.05)	.85 - .88	12.89%
	Adolescent	11	2717	.85 (.05)	.82 - .88	11.94%
Type of Sample:	Nonclinical	48	11321	.85 (.05)	.84 - .87	11.08%
	Clinical: Eating Disorder	4	423	.90 (.02)	.89 - .92	100.00%
	Mixed clinical and nonclinical	2	219	.88 (.05)	.81 - .95	20.57%
Test Language:	English	44	9077	.87 (.05)	.85 - .88	11.30%
	Non-English	10	2886	.81 (.04)	.79 - .84	25.69%
EAT-26 Dieting Factor		**24**	**10924**	**.80 (.07)**	**.77 - .83**	**5.50%**
Gender:	Female	22	10693	.80 (.07)	.77 - .83	5.06%
	Male	2	231	.81 (.03)	.77 - .85	100.00%
Age:	Adult	17	3870	.87 (.04)	.86 - .89	19.92%
	Adolescent	7	7054	.77 (.05)	.72 - .81	5.74%
Type of Sample:	Nonclinical	22	10602	.80 (.07)	.77 - .83	5.27%
Test Language:	English	21	10150	.80 (.07)	.77 - .83	5.35%
	Non-English	3	774	.86 (.04)	.81 - .90	17.23%
EAT-26 Bulimia and Food Preoccupation Factor		**23**	**10751**	**.67 (.11)**	**.63 - .71**	**5.44%**
Gender:	Female	21	10520	.67 (.11)	.62 - .72	5.12%
	Male	2	231	.71 (.11)	.56 - .87	17.26%
Age:	Adult	17	3870	.79 (.07)	.75 - .82	10.99%
	Adolescent	6	6881	.60 (.06)	.56 - .65	9.70%
Type of Sample:	Nonclinical	21	10429	.67 (.11)	.62 - .71	5.29%
Test Language:	English	20	9977	.67 (.11)	.62 - .72	4.92%
	Non-English	3	774	.71 (.04)	.67 - .76	60.81%

Table 4 Continued

Study Characteristics		*Reliability*				
		# of Data Points K	**Total Sample Size N**	**Sample-weighted Mean Alpha (SD)**	**95% CI**	**PVA_{SE}**
EAT-26 Oral Control Factor		**18**	**4475**	**.56 (.10)**	**.51 - .60**	**21.02%**
Gender:	Female	16	4244	.56 (.10)	.51 - .61	19.03%
	Male	2	231	.60 (.02)	.57 - .63	100.00%
Age:	Adult	14	3065	.56 (.11)	.50 - .61	19.04%
	Adolescent	4	1410	.57 (.06)	.51 - .63	34.67%
Type of Sample:	Nonclinical	16	4153	.55 (.08)	.51 - .59	27.63%
Test Language:	English	15	3701	.54 (.10)	.50 - .59	21.52%
	Non-English	3	774	.63 (.04)	.59 - .67	100.00%

Note. CI = Confidence Interval. PVA_{SE} = Percent of variance accounted for by sampling error.

greater than the female samples, and the adolescent samples had a similar mean estimate of internal consistency as compared to the adult samples. The non-English language test samples had higher reliability than the English samples.

For the EAT, the percentage of variance explained by sampling error ranged widely from approximately 5% to 100%. For the majority of the analyses, these values were less than 20%.

Bulimia Test

In 8 (19.51%) out of 41 studies, the researchers did not provide test score reliability for the BULIT, and in 16 (39.02%) studies, the researchers cited previous studies for score reliability. The researchers reported score reliability information for their sample in 17 (41.46%) studies but one study was excluded because the reliability coefficient was reported as a range of coefficients. The mean estimate of internal consistency for scores on the BULIT-R was .93, which was the highest mean for any of

the measures. Mean estimates of internal consistency for the BULIT are presented in Table 5.

For the BULIT-R, the female samples displayed greater reliability than the male samples. The adult samples had a higher estimate of internal consistency than the adolescent samples, and the mixed clinical and nonclinical samples displayed higher reliability than the nonclinical samples. All studies that reported internal consistency estimates used the English language version of the BULIT-R. For the BULIT-R, the percentage of variance accounted for by sampling error was slightly more restricted than for the other measures, with a range from approximately 12% to 63%. Again, the majority of these values were less than 20%.

Table 5

Mean Internal Consistency Estimates for the Bulimia Test-Revised (BULIT-R)

Study Characteristics		***Reliability***				
		# of Data Points K	**Total Sample Size N**	**Sample-weighted Mean Alpha (SD)**	**95% CI**	**PVA_{SE}**
BULIT-R		**16**	**3245**	**.93 (.03)**	**.92 - .94**	**12.68%**
Gender:	Female	13	2612	.94 (.02)	.93 - .95	17.90%
	Male	3	633	.90 (.03)	.87 - .93	25.56%
Age:	Adult	11	1830	.94 (.02)	.93 - .95	18.00%
	Adolescent	5	1415	.92 (.03)	.90 - .94	11.80%
Type of Sample:	Nonclinical	14	2918	.93 (.02)	.91 - .94	18.22%
	Mixed clinical and nonclinical	2	327	.97 (.005)	.97 - .98	63.07%

Note. CI = Confidence Interval. PVA_{SE} = Percent of variance accounted for by sampling error.

Test-Retest Reliability Analyses

Table 6 presents sample-weighted mean estimates of test-retest reliability for the EDI, EAT-26, and BULIT-R. For the EDI, the mean test-retest score reliability was .81, and for the EDI subscales, mean test-retest reliability estimates ranged from .42 to .77. The lowest mean test-retest reliability estimates were for the Maturity Fears (.42) and Interpersonal Distrust (.59) subscales. For the EAT-26, the mean test-retest reliability estimate was .87 and it was .85 for the BULIT-R. Due to the low number of data points available for each scale or subscale, these results should be interpreted with caution. Also, due to the low number of data points per measure, I did not conduct further analyses on these estimates, such as examining the effect of gender or retest interval.

Table 6

Mean Test-Retest Reliability Estimates for the EDI, EAT-26, and BULIT-R

Study Characteristics	***Reliability***				
	# of Data Points K	**Total Sample Size N**	**Sample-weighted Mean Reliability (SD)**	**95% CI**	**PVA_{SE}**
EDI Total	2	471	.81 (.06)	.72 - .90	12.68%
EDI Body Dissatisfaction Subscale	3	256	.77 (.14)	.62 - .92	10.66%
EDI Drive for Thinness Subscale	4	715	.60 (.15)	.45 - .75	10.07%
EDI Bulimia Subscale	4	715	.64 (.12)	.53 - .76	14.12%
EDI Perfectionism Subscale	3	695	.63 (.11)	.51 - .76	13.08%
EDI Interoceptive Awareness Subscale	2	236	.60 (.16)	.38 - .83	13.50%
EDI Ineffectiveness Subscale	2	236	.71 (.09)	.58 - .84	25.28%
EDI Maturity Fears Subscale	3	695	.42 (.11)	.29 - .55	22.68%
EDI Interpersonal Distrust Subscale	3	695	.59 (.10)	.47 - .70	19.51%
EAT-26	4	920	.87 (.02)	.85 - .88	100.00%
BULIT-R	2	341	.85 (.01)	.84 - .86	100.00%

Note. CI = Confidence Interval. PVA_{SE} = Percent of variance accounted for by sampling error

DISCUSSION

This study used reliability generalization procedures to find the mean score reliability for the EDI, the EAT, and the BULIT and to examine study characteristics that explain the variation in score reliability across studies. The reporting of score reliability information for the measures was higher (28.67% - 41.46%) than the reporting rate for other reliability generalization studies, such as the BDI (7.5%; Yin & Fan, 2000) and the NEO (15.2%; Caruso, 2000). However, given that score reliability information should be reported every time a measure is used, it is disappointing that approximately half of the studies using the EDI and the EAT failed to provide score reliability information.

Overall, mean reliability estimates for the measures were acceptable, with only the Asceticism subscale of the EDI and the Oral Control factor on the EAT-26 having questionable mean internal consistencies of .52 and .56, respectively. Except for scores on the Asceticism subscale, which had the lowest score reliability of all the EDI subscales, scores on the subscales added to the EDI when it was revised showed levels of reliability that were similar to the original 8 subscales. These findings suggest that there is not a clear advantage to using one of these measures over another based on mean score reliability, as all the measures generally had acceptable reliability overall and among subgroups of participants, such as clinical and nonclinical samples.

For the EDI, the majority of the subscales that measure specific eating disorder attitudes and behaviors, such as Bulimia and Perfectionism, displayed higher score reliability in clinical eating disorder samples than in nonclinical samples. This difference was not observed in the Drive for Thinness and Impulse Regulation subscales, and the

Body Dissatisfaction subscale showed similar score reliability in the two groups. One potential explanation for this discrepancy is that the attitudes measured by the Body Dissatisfaction and Drive for Thinness subscales are common in both nonclinical and eating disorder samples, contributing to more reliable measurement of these attitudes across sample type. The Perfectionism and Maturity Fears subscales displayed mean reliability above .70 in clinical eating disorder samples, but in the nonclinical samples, mean score reliability was below .70. It is more difficult to characterize the effect of sample type on the EAT and BULIT-R because scores on these measures were primarily reported for nonclinical samples.

Regarding participant age, mean score reliability tended to be slightly higher in the adult samples than in adolescent samples for all measures; however, reliability was generally acceptable in both groups. Again, the EDI Perfectionism and Maturity Fears subscale scores, as well as the EAT-26 Bulimia and Food Preoccupation subscale score, displayed mean reliability above .70 in the adult group but reliability below .70 in the adolescent group. The higher reliability in the adult group may be expected as the measures were developed in adult samples.

For all measures, there was a tendency for score reliability to be slightly higher in female samples than in the male samples, perhaps because eating disorder attitudes and behaviors are more prevalent in females resulting in greater score variability for this subpopulation.

Test-retest reliability analyses indicate that this type of reliability was generally acceptable for all three measures, with the lowest estimate found for the Maturity Fears

subscale of the EDI. Given the small amount of data available regarding test-retest reliability, these findings may be less statistically meaningful than the findings for internal consistency reliability estimates.

Although mean reliability estimates for scores on the EDI, EAT, and BULIT and their subscales were generally acceptable, the data indicate that some of the subscales display greater score reliability in certain subpopulations. It is important that researchers measure internal consistency for their sample every time a measure is used as characteristics of the sample affect test score reliability. This study has demonstrated that reliability estimates do not remain constant across studies; therefore, researchers should make sure that the scores for their sample are found to be reliable as an initial step in any study. Examining and reporting test score reliability should be included as descriptive information about the data. Additionally, researchers can tailor future studies to maximize score reliability, which is one factor that influences effect sizes.

One limitation of this study is the small number of data points available for some analyses. Although I reported information for many analyses where only 2-4 internal consistency estimates were available, these findings are less stable than a mean estimate based on 30-50 data points and therefore should be cautiously interpreted and were presented here only for the sake of completeness. Another potential limitation is having only one coder for the study. This decision was made because, unlike in traditional meta-analysis, the coder did not have to calculate effect sizes or other statistics and was only recording information as reported in articles; however, there is always the possibility that two coders could have disagreed about some of this basic information.

Implications

This study indicates that test score reliability for the EDI, EAT, and BULIT-R is greater for adult and clinical samples than for adolescent and nonclinical samples. Although it is important that disordered eating be reliably measured in an adult, clinical population, these findings are potentially troubling as it is also important that these concepts can be reliably measured in nonclinical adolescents who are at high risk for developing disordered eating attitudes and behaviors. In some cases, the differences in score reliability between adult and adolescent samples were small, and mean score reliability for adolescent samples remained acceptable overall. The differences between clinical eating disorder and nonclinical samples were generally larger. Without reliable measurement of these concepts in an at-risk adolescent population, researchers will have difficulty determining the true effectiveness of prevention programs designed to avoid or reduce future symptoms of eating disorders. Therefore, it is important for researchers to assess and report test score reliability with the measures they are using to determine the effectiveness of their programs.

Recommendations for Future RG Studies

The greatest difficulty to conducting this study was the lack of score reliability information reported in journal articles. With 50% of the coded studies not even mentioning score reliability of the measures, the ability to find meaningfully interpretable results was hindered as some analyses were conducted with a very small number of data samples. Potential explanations for this lack of reporting of score reliability estimates are that researchers are unaware of the importance of this

information so they do not calculate estimates of score reliability or that this information if being removed from manuscripts due to space limitations. Hopefully, as researchers, editors, and reviewers become more aware of the importance of score reliability information the reporting of these estimates will increase. One guaranteed way for reporting rates to increase is for journal editors to require this information be assessed and reported in order for a study to be published. This requirement would not only make it easier for future RG studies to be conducted but would also improve the quality of published studies.

CONCLUSIONS

The three measures examined in this study, the EDI, the EAT, and the BULIT-R generally had acceptable mean reliability estimates, with the BULIT-R displaying the highest mean score reliability. For all measures, score reliability tended to be slightly higher in adult and female samples. For the EDI, mean score reliability tended to be higher in clinical groups than in nonclinical groups for those scales specifically measuring eating disorder attitudes and behaviors, such as Bulimia and Perfectionism. For the EAT and the BULIT-R, mean score reliability also tended to higher in clinical samples compared to nonclinical groups; however, the effect for these two measures was more difficult to characterize because of the small number of studies using clinical participants. As the first reliability generalization analysis of any measure of eating disorder symptoms, this study advances the field by providing information about the typical score reliability of these measures and by illustrating factors that influence score reliability on these measures. This study also emphasizes the importance of researchers evaluating and reporting internal consistency every time a self-report measure is used.

REFERENCES

Agras, W. S. (2001). The consequences and costs of the eating disorders. *Psychiatric Clinics of North America, 24,* 371-379.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Arthur, W., Jr., Bennett, W., Jr., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS.* Mahwah, NJ: Lawrence Erlbaum Associates.

Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantitive issues are linked inextricably. *Educational and Psychological Measurement, 62,* 254-263.

Beck, A. T., & Steer, R. A. (1984). Internal consistencies of the original and revised Beck Depression Inventory. *Journal of Clinical Psychology, 40,* 1365-1367.

Caruso, J. C. (2000). Reliability generalization of the NEO Personality Scales. *Educational and Psychological Measurement, 60,* 236-254.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 197-334.

Dimitrov, D. M. (2002). Reliability: Arguments for multiple perspectives and potential problems with generalization across studies. *Educational and Psychological Measurement, 62,* 783-801.

Eason, S. (1991). Why generalizability theory yields better results than classical test theory: A primer with concrete examples. In B. Thompson (Ed.), *Advances in educational research: Substantive findings, methodological developments* (Vol. 1., pp. 83-98). Greenwich, CT: JAI Press.

Eberenz, K. P., & Gleaves, D. H. (1994). An examination of the internal consistency and factor structure of the Eating Disorder Inventory-2 in a clinical sample. *International Journal of Eating Disorders, 16,* 371-379.

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. B., & Cooper, Z. (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 Press.

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

Garner, D. M. (1995). Measurement of eating disorder psychopathology. In K. D. Brownell & C. G. Fairburn (Eds.), *Eating disorders and obesity: A comprehensive handbook* (pp. 117-121). New York: Guilford Press.

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

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

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

Gross, J., Rosen, J. C., Leitenberg, H., & Willmuth, M. E. (1986). Validity of the Eating Attitudes Test and the Eating Disorders Inventory in bulimia nervosa. *Journal of Consulting and Clinical Psychology, 54,* 875-876.

Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting "reliability generalization" studies. *Measurement and Evaluation in Counseling and Development, 35,* 113-126.

Hoek, H. W., & van Hoeken, D. (2003). Review of the prevalence and incidence of eating disorders. *International Journal of Eating Disorders, 34,* 383-396.

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2003). Reliability methods: A note on the frequency of use of various types. In B. Thompson (Ed.), *Score reliability: Contemporary thinking on reliability issues.* Thousand Oaks, CA: Sage Publications.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage Publications.

Hunter, J. E., & Schmidt, F. L. (1994). Correcting for sources of artificial variation across studies. In H. Cooper and L. V. Hedge (Eds.), *The handbook of research synthesis* (pp. 323-336). New York: Russell Sage Foundation.

Joiner, T. E., & Heatherton, T. F. (1998). First- and second-order factor structure of five subscales of the Eating Disorders Inventory. *International Journal of Eating Disorders, 23,* 189-198.

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.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales in counseling psychology research. *Journal of Counseling Psychology, 37,* 113-115.

Pomeroy, C., & Mitchell, J. E. (2002). Medical complications of anorexia nervosa and bulimia nervosa. In C. G. Fairburn & K. D. Brownell (Eds.), *Eating disorders and obesity: A comprehensive handbook* (2nd ed., pp. 278-285). New York: Guilford Press.

Raciti, M. C., & Norcross, J. C. (1987). The EAT and EDI: Screening, interrelationships, and psychometrics. *International Journal of Eating Disorders, 6,* 579-586.

Scheinberg, Z., Koslowsky, M., Bleich, A., Mark, M., Apter, A., Danon, Y., Solomon, Z., & Babur, I. (1993). Sensitivity, specificity, and positive predictive value as

measures of prediction accuracy: The case of the EAT-26. *Educational and Psychological Measurement, 53,* 831-839.

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

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

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

Thompson, B. (1994). Guidelines for authors reporting score reliability estimates. *Educational and Psychological Measurement, 54,* 837-847.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *The Journal of Experimental Education, 70,* 80-93.

Thompson, B. (2003). Understanding reliability and coefficient alpha, really. In B. Thompson (Ed.), *Score reliability: Contemporary thinking on reliability issues.* Thousand Oaks, CA: Sage Publications.

Thompson, J. K. (2004). Eating disorders and obesity: Definitions, prevalence, and associated features. In J. K. Thompson (Ed.), *Handbook of eating disorders and obesity.* Hoboken, NJ: Wiley.

Thurfjell, B., Edlund, B., Arinell, H., Hägglöf, B., Garner, D. M., & Engström, I. (2004). Eating Disorder Inventory for Children (EDI-C): Effects of age and gender in a Swedish sample. *European Eating Disorders Review, 12,* 256-264.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58,* 6-20.

Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62,* 562-569.

Viswesvaran, C., & Ones, D. S. (2000). Measurement error in "Big Five factors" personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement, 60,* 224-235.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.

Williamson, D. A., Anderson, D. A., & Gleaves, D. H. (1996). Anorexia nervosa and bulimia nervosa: Structured interview methodologies and psychological assessment. In J. K. Thompson (Ed.), *Body image, eating disorders, and obesity: An integrative guide for assessment and treatment* (pp. 205-223). Washington, DC: American Psychological Association.

Williamson, D. A., Womble, L. G., Smeets, M. A. M., Netemeyer, R. G., Thaw, J. M., Kutlesic, V., & Gleaves, D. H. (2002). Latent structure of eating disorder

symptoms: A factor analytic and taxometric investigation. *American Journal of Psychiatry, 159,* 412-418.

Yin, P., & Fan, X. (2000). Assessing the reliability of Beck Depression Inventory scores: Reliability generalization across studies. *Educational & Psychological Measurement, 60,* 201-223.

APPENDIX

REFERENCES INCLUDED IN RELIABILITY GENERALIZATION ANALYSES

Aarnia, K., & Lindeman, M. (2004). Magical food and health beliefs: A portrait of believers and functions of the beliefs. *Appetite, 43,* 65-74.

Adkins, E. C., & Keel, P. K. (2005). Does "excessive" or "compulsive" best describe exercise as a symptom of bulimia nervosa? *International Journal of Eating Disorders, 38,* 24-29.

Banasiak, S. J., Wertheim, E. H., Koemer, J., & Voudouris, N. J. (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.

Baş, M., Aşçı, H., Karabudak, E., & Kızıltan, G. (2004). Eating attitudes and their psychological correlates among Turkish adolescents. *Adolescence, 39,* 593-599.

Birkeland, R., Thompson, J. K., & Phares, V. (2005). Adolescent motherhood and postpartum depression. *Journal of Clinical Child and Adolescent Psychology, 34,* 292-300.

Bittinger, J. N., & Smith, J. E. (2003). Mediating and moderating effects of stress perception and situation type on coping responses in women with disordered eating. *Eating Behaviors, 4,* 89-106.

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.

Botta, R. A. (2003). For your health? The relationship between magazine reading and adolescents' body image and eating disturbances. *Sex Roles, 48,* 389-399.

Breen, H. B., & Espelage, D. L. (2004). Nutrition expertise in eating disorders. *Eating and Weight Disorders, 9,*120-125.

Brelsford, T. N., Hummel, R. M., & Barrios, B. A. (1992). The Bulimia Test–Revised: A psychometric investigation. *Psychological Assessment, 4,* 399-401.

Brookings, J. B., & Wilson, J. F. (1994). Personality and family-environment predictors of self-reported eating attitudes and behaviors. *Journal of Personality Assessment, 63,* 313-326.

Cash, T. F., & Hrabosky, J. I. (2003). The effects of psychoeducation and self-monitoring in a cognitive-behavioral program for body-image improvement. *Eating Disorders, 11,* 255-270.

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.

Compian, L., Gowen, L. K., & Hayward, C. (2004). Peripubertal girls' romantic and platonic involvement with boys: Associations with body image and depressive symptoms. *Journal of Research on Adolescence, 14,* 23-47.

Daubenmier, J. J. (2005). The relationship of yoga, body awareness, and body responsiveness to self-objectification and disordered eating. *Psychology of Women Quarterly, 29,* 207-219.

Davies, K., & Wardle, J. (1994). Body image and dieting in pregnancy. *Journal of*

Psychosomatic Research, 38, 787-799.

Davison, K. K., Markey, C. N., & Birch, L. L. (2003). A longitudinal examination of patterns in girls' weight concerns and body dissatisfaction from ages 5 to 9 years. *International Journal of Eating Disorders, 33,* 320-332.

Eberenz, K. P., & Gleaves, D. H. (1994). An examination of the internal consistency and factor structure of the Eating Disorder Inventory-2 in a clinical sample. *International Journal of Eating Disorders, 16,* 371-379.

Espelage, D. L, Mazzeo, S. E., Aggen, S. H., Quittner, A. L., Sherman, R., & Thompson, R. (2003). Examining the construct validity of the Eating Disorder Inventory. *Psychological Assessment, 15,* 71-80.

Espina Eizaguirre, A., de Cabezón, A. O. S., de Alda, I. O., Olariaga, L. J., & Juaniz, M. (2004). Alexithymia and its relationships with anxiety and depression in eating disorders. *Personality and Individual Differences, 36,* 321-331.

Favaro, A., Rodella, F. C., & Santonastaso, P. (2000). Binge eating and eating attitudes among Nazi concentration camp survivors. *Psychological Medicine, 30,* 463-466.

Fischer, S., Smith, G. T., Anderson, K. G., & Flory, K. (2003). Expectancy influences the operation of personality on behavior. *Psychology of Addictive Behaviors, 17,* 108-114.

Fitzgibbon, M. L., Sánchez-Johnsen, L. A., Martinovich, Z. (2003). A test of the continuity perspective across bulimic and binge eating pathology. *International Journal of Eating Disorders, 34,* 83-97.

Francis, L. A., & Birch, L. L. (2005). Maternal influences on daughters' restrained eating

behavior. *Health Psychology, 24,*548-554.

Franko, D. L., Striegel-Moore, R. H., Barton, B. A., Schumann, B. C., Garner, D. M., Daniels, S. R., Schreiber, G. B., & Crawford, P. B. (2004.) Measuring eating concerns in black and white adolescent girls. *International Journal of Eating Disorders, 35,* 179-189.

Fung, M. S. C., & Yuen, M. (2003). Body image and eating attitudes among adolescent Chinese girls in Hong Kong. *Perceptual and Motor Skills, 96,* 57-66.

Furnham, A., Badmin, N., & Sneade, I. (2002). Body image dissatisfaction: Gender differences in eating attitudes, self-esteem, and reasons for exercise. *The Journal of Psychology, 136,* 581-596.

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

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

Graber, J. A., Tyrka, A. R., & Brooks-Gunn, J. (2003). How similar are correlates of different subclinical eating problems and bulimia nervosa? *Journal of Child Psychology & Psychiatry, 44,* 262-273.

Haase, A. M., Prapavessis, H., & Owens, R. G. (2002). Perfectionsim, social physique anxiety and disordered eating: A comparison of male and female elite athletes. *Psychology of Sport and Exercise, 3,* 209-222.

Harrison, K. (2000). The body electric: Thin-ideal media and eating disorders in

adolescents. *Journal of Communication, 50,* 119-143.

Harrison, K., & Fredrickson, B. L. (2003). Women's sports media, self-objectification, and mental health in black and white adolescent females. *Journal of Communication, 53,* 216-232.

Hausenblas, H. A., Janelle, C. M., Gardner, R. E., & Focht, B. C. (2004). Viewing physique slides: Affective responses of women at high and low drive for thinness. *Journal of Social and Clinical Psychology, 23,* 45-60.

Holt, M. K., & Espelage, D. L. (2002). Problem-solving skills and relationship attributes among women with eating disorders. *Journal of Counseling and Development, 80,* 346-354.

Hopkinson, R. A., & Lock, J. (2004). Athletics, perfectionism, and disordered eating. *Eating and Weight Disorders, 9,* 99-106.

Humphry, T. A., & Ricciardelli, L. A. (2004). The development of eating pathology in Chinese-Australian women: Acculturation versus culture clash.

Hund, A. R., & Espelage, D. L. (2005). Childhood sexual abuse, disordered eating, alexithymia, and general distress: A mediation model. *Journal of Counseling Psychology, 52,* 559-573.

Iyer, D. S., & Haslam, N. (2003). Body image and eating disturbance among south Asian-American women: The role of racial teasing. *International Journal of Eating Disorders, 34,* 142-147.

Jackson, T., Weiss, K. E., Lunquist, J. J., & Soderlind, A. (2005). Sociotropy and perceptions of interpersonal relationships as predictors of eating disturbances

among college women: Two prospective studies. *The Journal of Genetic Psychology, 166,* 346-359.

Janelle, C. M., Hausenblas, H. A., Fallon, E. A., & Gardner, R. E. (2003). A visual search examination of attentional biases among individuals with high and low drive for thinness. *Eating and Weight Disorders, 8,* 138-144.

Johnson, C. E., & Petrie, T. A. (1996). Relationship of gender discrepancy to psychological correlates of disordered eating in female undergraduates. *Journal of Counseling Psychology, 43,* 473-479.

Johnson, C. S., & Bedford, J. (2004). Eating attitudes across age and gender groups: A Canadian study. *Eating and Weight Disorders, 9,* 16-23.

Joiner, G. W., & Kashubeck, S. (1996). Acculturation, body image, self-esteem, and eating-disorder symptomatology in adolescent Mexican American women. *Psychology of Women Quarterly, 20,* 419-435.

Joiner, T. E., Jr., Heatherton, T. F., & Keel, P. K. (1997). Ten-year stability and predictive validity of five bulimia-related indicators. *American Journal of Psychiatry, 154,* 1133-1138.

Jordan, P. J., Redding, C. A., Troop, N. A., Treasure, J., & Serpell, L. (2003). Developing a stage of change measure for assessing recovery from anorexia nervosa. *Eating Behaviors, 3,* 365-385.

Keery, H., Boutelle, K., van den Berg, P., & Thompson, J. K. (2005). The impact of appearance-related teasing by family members. *Journal of Adolescent Health, 37,* 120-127.

Kirk, G., Singh, K., & Getz, H. (2001). Risk of eating disorders among female college athletes and nonathletes. *Journal of College Counseling, 4,* 122-132.

Koslowsky, M., Scheinberg, Z., Bleich, A., Mark, M., Apter, A., Danon, Y., & Solomon, Z. (1992). The factor structure and criterion validity of the short form of the Eating Attitudes Test. *Journal of Personality Assessment, 58,* 27-35.

Lee, S., Kwok, K., Liau, C., & Leung, T. (2002). Screening Chinese patients with eating disorders using the Eating Attitudes Test in Hong Kong. *International Journal of Eating Disorders, 32,* 91-97.

Lee, S., & Lee, A. M. (2000). Disordered eating in three communities of China: A comparative study of female high school students in Hong Kong, Shenzhen, and rural Hunan. *International Journal of Eating Disorders, 27,* 317-327.

Lester, R., & Petrie, T. A. (1995). Personality and physical correlates of bulimic symptomatology among Mexican American female college students. *Journal of Counseling Psychology, 42,* 199-203.

Leung, F., Wang, J., & Tang, C. W. (2004). Psychometric properties and normative data of the Eating Disorder Inventory among 12 to 18 year old Chinese girls in Hong Kong. *Journal of Psychosomatic Research, 57,* 59-66.

Lorenzo, C. R., Lavori, P. W., & Lock, J. D. (2002). Eating attitudes in high school students in the Philippines: A preliminary study. *Eating and Weight Disorders, 7,* 202-209.

Martin, K. A., & Hausenblas, H. A. (1998). Psychological commitment to exercise and eating disorder symptomatology among female aerobic instructors. *The Sport*

Psychologist, 12, 180-190.

Mazzeo, S. E. (1999). Modification of an existing measure of body image preoccupation and its relationship to disordered eating in female college students. *Journal of Counseling Psychology, 46,* 42-50.

McVey, G. L., Davis, R., Tweed, S., & Shaw, B. F. (2004). Evaluation of a school-based program designed to improve body image satisfaction, global self-esteem, and eating attitudes and behaviors: A replication study. *International Journal of Eating Disorders, 36,* 1-11.

McVey, G. L., Pepler, D., Davis, R., Flett, G. L., & Abdolell, M. (2002). Risk and protective factors associated with disordered eating during early adolescence. *Journal of Early Adolescence, 22,* 75-95.

Miotto, P., De Coppi, M., Frezza, M., & Preti, A. (2003). The spectrum of eating disorders: Prevalence in an area of Northeast Italy. *Psychiatry Research, 119,* 145-154.

Moradi, B., Dirks, D., & Matteson, A. V. (2005). Roles of sexual objectification experiences and internalization of standards of beauty in eating disorder symptomatology: A test and extension of objectification theory. *Journal of Counseling Psychology, 52,* 420-428.

Nobakht, M., & Kezhkam, M. (2000). An epidemiological study of eating disorders in Iran. *International Journal of Eating Disorders, 28,* 265-271.

Ohring, R., Graber, J. A., Brooks-Gunn, J. (2002). Girls' recurrent and concurrent body dissatisfaction: Correlates and consequences over 8 years. *International Journal*

of Eating Disorders, 31, 404-415.

Pelletier, L. G., Dion, S., & Lévesque, C. (2004). Can self-determination help protect women against sociocultural influences about body image and reduce their risk of experiencing bulimic symptoms? *Journal of Social and Clinical Psychology, 23,* 61-88.

Piran, N., & Cormier, H. C. (2005). The social construction of women and disordered eating patterns. *Journal of Counseling Psychology, 52,* 549-558.

Podar, I., Hannus, A., & Allik, J. (1999). Personality and affectivity characteristics associated with eating disorders: A comparison of eating disordered, weight-preoccupied, and normal samples. *Journal of Personality Assessment, 73,* 133-147.

Prouty, A. M., Protinsky, H. O., & Canady, D. (2002). College women: Eating behaviors and help-seeking preferences. *Adolescence, 37,* 353-363.

Raciti, M. C., & Norcross, J. C. (1987). The EAT and EDI: Screening, interrelationships, and psychometrics. *International Journal of Eating Disorders, 6,* 579-586.

Rhea, D. J. (1999). Eating disorder behaviors of ethnically diverse urban female adolescent athletes and non-athletes. *Journal of Adolescence, 22,* 379-388.

Rizvi, S. L, Stice, E., & Agras, W. S. (1999). Natural history of disordered eating attitudes and behaviors over a 6-year period. *International Journal of Eating Disorders, 26,* 406-413.

Robinson, T. N., Killen, J. D., Litt, I. F., Hammer, L. D., Wilson, D. M., Haydel, K. F., Hayward, C., & Taylor, C. B. (1996). Ethnicity and body dissatisfaction: Are

Hispanic and Asian girls at increased risk for eating disorders? *Journal of Adolescent Health, 19,* 384-393.

Russell, C. J., & Keel, P. K. (2002). Homosexuality as a specific risk factor for eating disorders in men. *International Journal of Eating Disorders, 31,* 300-306.

Ryu, H. R., Lyle, R. M., Galer-Unti, R. A., & Black, D. R. (1999). Cross-cultural assessment of eating disorders: Psychometric characteristics of a Korean version of the Eating Disorder Inventory-2 and the Bulimia Test-Revised. *Eating Disorders, 7,* 109-122.

Santonastaso, P., Mondini, S., & Favaro, A. (2002). Are fashion models a group at risk for eating disorders and substance abuse? *Psychotherapy and Psychosomatics, 71,* 168-172.

Scheinberg, Z., Koslowsky, M., Bleich, A., Mark, M., Apter, A., Danon, Y., Solomon, Z., & Babur, I. (1993). Sensitivity, specificity, and positive predictive value as measures of prediction accuracy: The case of the EAT-26. *Educational and Psychological Measurement, 53,* 831-839.

Schoemaker, C., van Strien, T., & van der Staak, C. (1994). Validation of the Eating Disorders Inventory in a nonclinical population using transformed and untransformed responses. *International Journal of Eating Disorders, 15,* 387-393.

Sherry, S. B., Hewitt, P. L., Besser, A., McGee, B. J., & Flett, G. L. (2004). Self-oriented and socially prescribed perfectionism in the Eating Disorder Inventory Perfectionism subscale. *International Journal of Eating Disorders, 35,* 69-79.

Shore, R. A., & Porter, J. E. (1990). Normative and reliability data for 11 to 18 year olds on the Eating Disorder Inventory. *International Journal of Eating Disorders, 9,* 201-207.

Siervo, M., Boschi, V., Papa, A., Bellini, O., & Falconi, C. (2005). Application of the SCOFF, Eating Attitude Test 26 (EAT 26) and Eating Inventory (TFEQ) questionnaires in young women seeking diet-therapy. *Eating and Weight Disorders, 10,* 76-82.

Sim, L., & Zeman, J. (2004). Emotion awareness and identification skills in adolescent girls with bulimia nervosa. *Journal of Clinical Child and Adolescent Psychology, 33,* 760-771.

Slater, A., & Tiggemann, M. (2002). A test of objectification theory in adolescent girls. *Sex Roles, 46,* 343-349.

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

Spillane, N. S., Boerner, L. M., Anderson, K. G., & Smith, G. T. (2004). Comparability of the Eating Disorder Inventory-2 between women and men. *Assessment, 11,* 85-93.

Stice, E., Akutagawa, D., Gaggar, A., & Agras, W. S. (2000). Negative affect moderates the relation between dieting and binge eating. *International Journal of Eating Disorders, 27,* 218-229.

Stice, E., Mazotti, L., Krebs, M., & Martin, S. (1998). Predictors of adolescent dieting behaviors: A longitudinal study. *Psychology of Addictive Behaviors, 12,* 195-205.

Tachikawa, H., Yamaguchi, N., Haianaka, K., Kobayashi, J., Sato, S., Mizukami, K., Asada, T., & Sugie, M. (2004). The Eating Disorder Inventory-2 in Japanese clinical and non-clinical samples: Psychometric properties and cross-cultural implications. *Eating and Weight Disorders, 9,* 107-113.

Tasca, G. A., Illing, V., Lybanon-Daigle, V., Bissada, H., & Balfour, L. (2003). Psychometric properties of the Eating Disorders Inventory-2 among women seeking treatment for binge eating disorder. *Assessment, 10,* 228-236.

Tchanturia, K., Katzman, M., Troop, N. A., & Treasure, J. (2002). An exploration of eating disorders in a Georgian sample. *International Journal of Social Psychiatry, 48,* 220-230.

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

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

Thome, J., & Espelage, D. L. (2004). Relations among exercise, coping, disordered eating, and psychological health among college students. *Eating Behaviors, 5,* 337-351.

Thurfjell, B., Edlund, B., Arinell, H., Hägglöf, B., & Engström, 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.

Tiggemann, M. (2005). Television and adolescent body image: The role of program content and viewing motivation. *Journal of Social and Clinical Psychology, 24,* 361-381.

Tylka, T. L. (2004). The relation between body dissatisfaction and eating disorder symptomatology: An analysis of moderating variables. *Journal of Counseling Psychology, 51,* 178-191.

Tylka, T. L., & Subich, L. M. (2002). A preliminary investigation of the eating disorder continuum with men. *Journal of Counseling Psychology, 49,* 273-279.

Tylka, T. L., & Subich, L. M. (2003). Revisiting the latent structure of eating disorders: Taxometric analyses with nonbehavioral indicators. *Journal of Counseling Psychology, 50,* 276-286.

van den Berg, P., Thompson, J. K., Obremski-Brandon, K., & Coovert, M. (2002). The Tripartite Influence model of body image and eating disturbance. A covariance structure modeling investigation testing the mediational role of appearance comparison. *Journal of Psychosomatic Research, 53,* 1007-1020.

van Strien, T., & Ouwens, M. (2003). Validation of the Dutch EDI-2 in one clinical and two nonclinical populations. *European Journal of Psychological Assessment, 19,* 66-84.

Vander Wal, J. S. (2004). Eating and body image concerns among average-weight and obese African American and Hispanic girls. *Eating Behaviors, 5,* 181-187.

Vanderheyden, D. A., Fekken, G. C., & Boland, F. J. (1988). Critical variables associated with bingeing and bulimia in a university population: A factor analytic

study. *International Journal of Eating Disorders, 7,* 321-329.

Vincent, M. A., McCabe, M. P., & Ricciardelli, L. A. (1999). Factorial validity of the Bulimia Test-Revised in adolescent boys and girls. *Behaviour Research and Therapy, 37,* 1129-1140.

Vohs, K. D., Bardone, A. M., Joiner, T. E., Jr., Abramson, L. Y., & Heatherton, T. F. (1999). Perfectionism, perceived weight status, and self-esteem interact to predict bulimic symptoms: A model of bulimic symptom development. *Journal of Abnormal Psychology, 108,* 695-700.

Wardle, J., & Watters, R. (2004). Sociocultural influences on attitudes to weight and eating: Results of a natural experiment. *International Journal of Eating Disorders, 35,* 589-596.

Wassenaar, D., le Grange, D., Winship, J., & Lachenicht, L. (2000). The prevalence of eating disorder pathology in a cross-ethnic population of female students in South Africa. *European Eating Disorders Review, 8,* 225-236.

Wear, R. W., & Pratz, O. (1987). Test-retest reliability for the Eating Disorder Inventory. *International Journal of Eating Disorders, 6,* 767-769.

Welch, G., & Hall, A. (1989). The reliability and discriminant validity of three potential measures of bulimic behaviours. *Journal of Psychiatric Research, 23,* 125-133.

Wichstrøm, L. (2000). Psychological and behavioral factors unpredictive of disordered eating: A prospective study of the general adolescent population in Norway. *International Journal of Eating Disorders, 28,* 33-42.

Wilhelmsson, M., & Andersson, A. L. (2005). An attempt at distinguishing subgroups of

women with anorexia nervosa and bulimia nervosa by means of the Defense Mechanism Technique modified (DMTm) and the Eating Disorder Inventory (EDI). *Eating and Weight Disorders, 10,* 175-186.

Williams, T. L., & Gleaves, D. H. (2003). Childhood sexual abuse, body image, and disordered eating: A structural modeling analysis. *Journal of Trauma & Dissociation, 4,* 91-108.

Wonderlich, A. L., Ackard, D. M., & Henderson, J. B. (2005). Childhood beauty pageant contestants: Associations with adult disordered eating and mental health. *Eating Disorders, 13,* 291-301.

Wood, A., Waller, G., Miller, J., & Slade, P. (1992). The development of Eating Attitude Test scores in adolescence. *International Journal of Eating Disorders, 11,* 279-282.

Zabinski, M. F., Calfas, K. J., Gehrman, C. A., Wilfley, D. A., & Sallis, J. F. (2001). Effects of a physical activity intervention on body image in university seniors: Project GRAD. *Annals of Behavioral Medicine, 23,* 247-252.

Zmijewski, C. F., & Howard, M. O. (2003). Exercise dependence and attitudes toward eating among young adults. *Eating Behaviors, 4,* 181-195.

VITA

Crystal Anne Pearson
Brooke Army Medical Center
Department of Behavioral Medicine
BLDG 3600, 3851 Roger Brooke Dr
Fort Sam Houston, TX 78234-6200
crystal.pearson@amedd.army.mil

Education

2006 Doctor of Philosophy, Texas A&M University, Clinical Psychology
2003 Master of Science, Texas A&M University, Clinical Psychology
1999 Bachelor of Arts, Mississippi State University, Psychology

Clinical Experience

2005-2006 Clinical Psychology Internship, Brooke Army Medical Center
2000-2004 Texas A&M Psychology Clinic
2000-2002 Federal Prison Camp, Bryan, TX

Publications

Pearson, C. A., & Gleaves, D. H. (2006). The multiple dimensions of perfectionism and their relation with eating disorder features. *Personality and Individual Differences, 41,* 225-235.

Fingeret, M. C., Gleaves, D. H., & Pearson, C. A. (2004). On the methodology of body image assessment: The use of figural rating scales to evaluate body dissatisfaction and the ideal body standards of women. *Body Image, 1,* 207-212.

Presentations *(selected)*

Pearson, C. A., & Gleaves, D. H. (2004). Examining study characteristics that affect the reliability of scores on the Eating Disorder Inventory: A reliability generalization study. Poster presented at the annual meeting of the Association for Advancement of Behavior Therapy, New Orleans, LA, November 2004.

Brown, J. D., Fingeret, M. C., Pearson, C. A., & Gleaves, D. H. (2003). The relationship between sociocultural appearance standards and body image for males and females. Poster presented at the annual meeting of the Association for Advancement of Behavior Therapy, Boston, MA, November 2003.

Pearson, C. A., Brown, J. D., & Gleaves, D. H. (2002). Examining the relationship between perfectionism, self-esteem, body satisfaction, and bulimic behavior. Poster presented at the annual meeting of the Association for Advancement of Behavior Therapy, Reno, NV, November 2002.

Brown, J. D., Gleaves, D. H., & Pearson, C. A. (2002). Male and female body image: Critical measurement issues. Poster presented at the annual meeting of the Association for Advancement of Behavior Therapy, Reno, NV, November 2002.