

RESEARCH ARTICLE

Open Access

Measuring eating disorder attitudes and behaviors: a reliability generalization study

David H Gleaves^{1*}, Crystal A Pearson², Suman Ambwani³ and Leslie C Morey²

Abstract

Background: Although score reliability is a sample-dependent characteristic, researchers often only report reliability estimates from previous studies as justification for employing particular questionnaires in their research. The present study followed reliability generalization procedures to determine the mean score reliability of the Eating Disorder Inventory and its most commonly employed subscales (Drive for Thinness, Bulimia, and Body Dissatisfaction) and the Eating Attitudes Test as a way to better identify those characteristics that might impact score reliability.

Methods: Published studies that used these measures were coded based on their reporting of reliability information and additional study characteristics that might influence score reliability.

Results: Score reliability estimates were included in 26.15% of studies using the EDI and 36.28% of studies using the EAT. Mean Cronbach's alphas for the EDI (total score = .91; subscales = .75 to .89), EAT-40 (total score = .81) and EAT-26 (total score = .86; subscales = .56 to .80) suggested variability in estimated internal consistency. Whereas some EDI subscales exhibited higher score reliability in clinical eating disorder samples than in nonclinical samples, other subscales did not exhibit these differences. Score reliability information for the EAT was primarily reported for nonclinical samples, making it difficult to characterize the effect of type of sample on these measures. However, there was a tendency for mean score reliability to be higher in the adult (vs. adolescent) samples and in female (vs. male) samples.

Conclusions: Overall, this study highlights the importance of assessing and reporting internal consistency during every test administration because reliability is affected by characteristics of the participants being examined.

Keywords: Eating disorders, Assessment, Psychometric properties, Reliability generalization, Eating disorder inventory, Eating attitudes test

Background

Although estimates for anorexia nervosa and bulimia nervosa approximate .3% and 1% respectively when using strict diagnostic criteria [1], disturbances in eating behavior and body image affect large numbers of individuals [2,3] and recent evidence suggests increases in the annual incidence of EDs in the U.K. [4] and a substantial rise in the point prevalence of ED behaviors in Australia [5]. Several measures are available for the assessment of ED symptomatology, but researchers or clinicians may falsely assume that these tools retain adequate psychometric properties such as internal consistency across all circumstances [6]. For instance,

while reporting research results, authors frequently refer to the "reliability of a test", a shorthand phrase that contributes to the misunderstanding by many researchers and students that tests, rather than scores, may be reliable [7]. This distinction between the reliability of test scores during a particular administration versus test reliability is significant; as emphasized by Wilkinson and the APA Task Force on Statistical Inference [8], "It is important to remember that a test is not reliable or unreliable. Reliability is a property of the scores on a test for a particular population of examinees" (p. 596) and thus reliability coefficients may vary depending on characteristics of the sample.

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,

* Correspondence: David.Gleaves@unisa.edu.au

¹School of Psychology, Social Work and Social Policy, University of South Australia, Magill Campus, GPO Box 2471, Adelaide, SA 5001, Australia
Full list of author information is available at the end of the article

leading to a potential problem with validity of the data [9]; reliability of test scores is viewed theoretically as a necessary condition to establish validity, as “unreliable scores measure nothing” ([9], p. 6). Second, poor score reliability may hinder the ability to find statistically, clinically, or practically significant effects [9]. When interpreting effect sizes, score reliability is an important factor to consider because measurement error impacts effect size [10-13], as a larger standard error contributes to a less precise effect size value [14]. Measurement errors cause observed effects to fluctuate across studies and may lead to underestimation of true effects [10]. This has led to recommendations for correcting effect size estimates for unreliable scores [15]. Third, total score variance affects reliability of the data set, and total score variance is impacted by characteristics of the participants [6,7]. Because score variability is a property of the data, reliability estimates will not remain constant across studies and should therefore be evaluated and reported as part of the process of describing the data.

Given the importance of test score reliability to scientific research, it is surprising that the editorial policies of journals often do not require this information to be reported and many authors do not report reliability estimates for their data [13,16]. Studies examining reporting rates for score reliability have estimates ranging from 7.5% for the Beck Depression Inventory [6] to 15.2% for the NEO Five-Factor Inventory [17], and Henson and Thompson [18] suggested that reporting rates are unlikely to exceed 40% for any test. Although reliability generalization studies have been conducted for self-report measures assessing various aspects of psychopathology, such as autism [19], substance abuse [20], depression [21], obsessive-compulsive symptoms [22] and general psychopathology [23], no published studies to date have evaluated the same for eating disorder symptoms.

Given the significance of accurately assessing test reliability, the present study employed reliability generalization (RG) procedures to report the mean score reliability for common measures of eating disorder symptoms and to examine variability in these estimates across sample characteristics. 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 [16]. RG is consistent with previous work on validity generalization [24], in which researchers conduct analyses to determine if the validity of scores on a test was generalizable to different samples [12]. 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 [6].

Two commonly used self-report measures of eating disordered attitudes and behaviors are the Eating Disorder Inventory (EDI; [25]) and the Eating Attitudes Test (EAT; [26]), both of which are available in revised forms (EDI-2; [27]; EAT-26; [28]). Research suggests that the EDI can distinguish individuals with AN [25] and BN [29] from nonclinical respondents. Conversely, the Eating Attitudes Test (EAT; [26]) assesses thoughts and behaviors related to anorexia nervosa and may be administered in the original 40-item version or a 26-item short form (EAT-26; [28]), both of which are typically highly correlated ($r = .98$; [30]). The EAT has also been shown to discriminate individuals with bulimia nervosa from control participants [29], eating disordered patients and controls, and binge eating patients from those with anorexia nervosa and bulimia nervosa [31].

An examination of the factors that influence reliability in eating disorder assessment can facilitate an understanding of how sample characteristics may contribute to variability in score quality. For example, other factors being equal, a heterogeneous set of participants will produce higher score reliability than a more homogenous group of participants [6]. Participant characteristics such as the type of sample, age and gender should be considered in evaluating score reliability. Other study factors potentially impacting test score reliability are sample size, type of reliability, culture/ethnicity, test format, test length, and test language. By identifying the conditions under which a test's scores display higher or lower reliability, researchers will be able to tailor future studies about eating disordered attitudes and behaviors to conditions that will maximize score reliability and thereby yield additional control over one factor that influences effect sizes. Thus, in the present study we used RG procedures to study the mean score reliability for different versions of the EDI and the EAT to explore how score reliability of these eating disorder measures varies across studies, and explore the study characteristics that account for this variation.

Methods

Procedure

The present study followed five steps for designing an RG study as recommended by Henson and Thompson [18]: selecting the measures to be analyzed, developing a coding sheet, collecting data, identifying potential dependent variables, and conducting analyses. For test selection, studies that utilized various forms of the EDI or the EAT were selected due to their common use as measures of eating disorder symptomatology in clinical and research settings. For developing a coding form and data collection, relevant reports were gathered through database searches of PsycINFO using the terms Eating Disorder Inventory and Eating Attitudes Test. The search

period was from when the tests were first published (1979 for the EAT and 1983 for the EDI) through to the end of 2005, so approximately 27 years and 23 years of research publications for the EAT and EDI, respectively. The searches resulted in 873 references for the EDI and 601 for the EAT during that time period. Included references were published empirical journal articles; books/book chapters, theoretical articles, review articles, case studies, dissertations, and meta-analyses, articles not published in English were excluded from this study. Based on these criteria, 283 studies of the EDI and 215 studies of the EAT were reviewed and coded (see Figure 1 for a flow chart). The data coding sheet included codes for whether or not 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 was used to code all studies because, unlike in more traditional meta-analyses, this study did not require any calculations to be made. Separate analyses were conducted using internal consistency and test-retest coefficients as dependent variables.

When using reliability estimates, some researchers combine Cronbach’s alpha [32] estimates with test-retest reliability estimates as a single dependent variable, but Dimitrov [33] cautioned against combining these estimates as they are not equivalent, and combining them together could lead to “mixing apples and oranges” ([33],

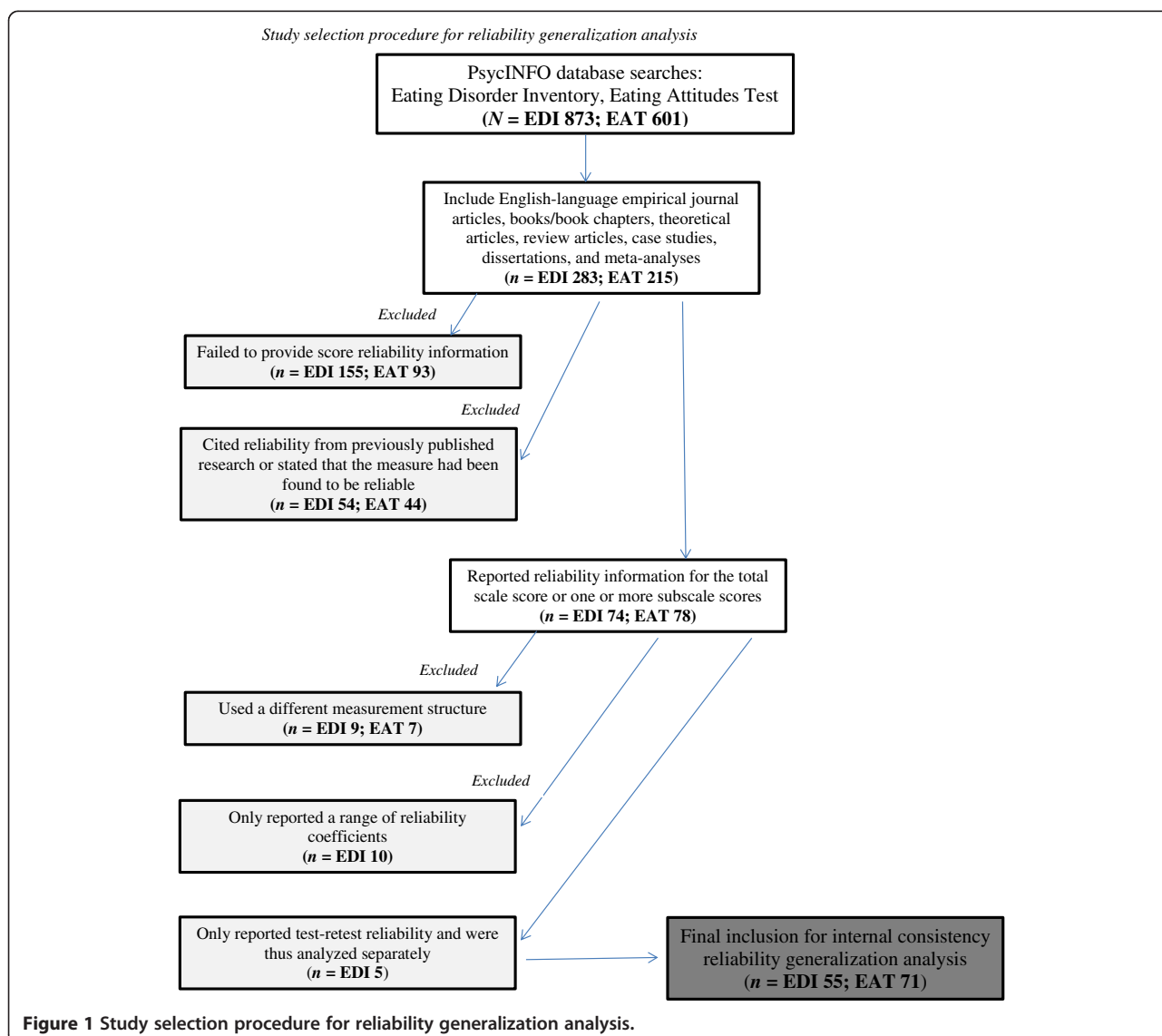


Figure 1 Study selection procedure for reliability generalization analysis.

p. 792). Thus, after coding all the studies, we 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 published guidelines for conducting a reliability generalization study [18] and for conducting a meta-analysis [34] were followed. 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 (54.77%) out of 283 studies that used the EDI or EDI-2, the researchers did not provide any score reliability information (see Figure 1). In 54 (19.08%) 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 74 (26.15%) 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 measurement structure (see [25,27,29,31,35-87] for included studies). 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.

Table 1 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. All of the coded study factors could not be analyzed 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. Mean estimates of internal consistency for scores on the subscales ranged from .75 to .89 and for scores on the EDI, the mean estimate was .90. No studies reported estimates of internal consistency for the total score on the EDI-2. The Bulimia subscale had higher score reliability in clinical eating disorder samples than in nonclinical samples, whereas the Drive for Thinness 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.

An examination of moderators of score reliability suggested that for the total EDI, there were some differences across adult vs. adolescent samples and clinical vs. nonclinical samples (see Table 1 for confidence intervals). For the most commonly employed EDI subscales (Body Dissatisfaction, Drive for Thinness, and Bulimia), mean estimates for internal consistency were .89, .85, and .75, respectively. For the Body Dissatisfaction subscale, the mean score reliability was higher in the female and mixed gender samples than in the male samples, and the adult samples had greater reliability than the adolescent samples. For the Drive for Thinness subscale, 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 language samples displayed greater reliability in their scores than the non-English test language samples. For the Bulimia subscale, 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 (see Table 1).

The percent of variance explained by sampling error varied widely for the EDI and its subscales, ranging from 1.65% 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

Reliability information was not provided in 93 (43.26%) of 215 studies utilizing the EAT, and in 44 (20.47%) studies, the researchers made some reference to score

Table 1 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 (<.01)	.93 - .93	100.00%
Type of sample: Nonclinical	4	2989	.90 (.06)	.84 - .96	1.24%
Clinical: Eating disorder	2	280	.95 (.01)	.93 - .96	99.32%
Test language: English	4	995	.85 (.08)	.77 - .93	4.35%
Non-English	2	2274	.93 (<.01)	.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%

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

reliability from previously published studies (see Figure 1 for a flow chart). Reliability information for either the total EAT score or one or more factor scores for their sample was reported 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 (see [26,28-30,69-135] for included studies). Results

indicate that 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 2 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.

An examination of prospective moderators suggested that for the EAT-40, the female samples had higher reliability than the male and mixed gender samples, and the mixed gender group displayed higher reliability than the male group. However, 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-speaking samples also had a higher mean estimate of internal consistency than the non-English samples. Regarding subscale scores, 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 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. Overall, 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%.

Test-retest reliability analyses

Table 3 presents sample-weighted mean estimates of test-retest reliability for the EDI and EAT-26. 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 Drive for Thinness [60] and Bulimia [64] subscales. For the EAT-26, the mean test-retest reliability estimate was .87. However, due to the low number of data points available for each scale

or subscale, these results should be interpreted with caution.

Discussion

This study used reliability generalization procedures to find the mean score reliability for different versions of the EDI and the EAT and to examine study characteristics (i.e., moderators) that might explain the variation in score reliability across studies. The reporting rate of score reliability information for the measures was higher (26.15% - 41.46%) than the reporting rate for other RG studies, such as the BDI (7.5%) [6] and the NEO (15.2%) [17]. However, given that score reliability information should be reported every time a measure is used, it is disappointing that such a large proportion of the studies using the EDI and the EAT failed to provide such information.

Overall, mean reliability estimates for the measures were acceptable, with only the Oral Control factor on the EAT-26 exhibiting questionable mean internal consistency [56]. For the EDI, the Bulimia subscale, which was designed to measure specific eating disorder attitudes and behaviors, displayed higher score reliability in clinical eating disorder samples than in nonclinical samples. Conversely, the Drive for Thinness and Body Dissatisfaction subscales exhibited similar score reliability in clinical and nonclinical 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. It is more difficult to characterize the effect of sample type on the two versions of the EAT because scores on these measures were primarily reported for nonclinical samples.

Regarding participant age and gender, 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. However, the EAT-26 Bulimia and Food Preoccupation subscale scores displayed mean reliability above .70 in the adult group but 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 among women resulting in greater score variability for this subpopulation.

Test-retest reliability analyses indicate that this type of reliability was generally acceptable for both measures, with the lowest estimate found for the EDI Drive for Thinness subscale, followed by the EDI Bulimia subscale. Although the EDI Bulimia scale did exhibit lower internal consistency estimates among certain samples, the

Table 2 Mean internal consistency estimates for the eating attitudes test (EAT)

Study characteristics	# of data points K	Total sample size N	Reliability		
			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%
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%

Table 2 Mean internal consistency estimates for the eating attitudes test (EAT) (Continued)

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.

EDI Drive for Thinness scale appeared to have similar score reliability across diverse samples; thus, it is somewhat surprising that this scale should exhibit the lowest overall test-retest reliability. One possibility for this might be that the construct assessed by this scale may be more subject to temporal fluctuations than some of the other constructs assessed by the EDI. Alternatively, the EDI Drive for Thinness scale is frequently employed in experimental studies involving brief interventions designed to change participants' attitudes toward thinness (e.g., [136]), suggesting that this tool may be sensitive to fluctuations in drive for thinness. Finally, 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 their subscales were generally acceptable, the data indicate that some of the subscales display greater score reliability in female, adult, and clinical (eating disorder) subpopulations, but with some variability across the different subscales. 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. The present study demonstrates that reliability estimates do not remain constant across studies; therefore, researchers should ensure 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 information was reported for 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. In addition to addressing some of these design limitations and conducting a more recent search of the literature, future research could also examine the predictors of reliability estimates for other frequently employed assessment tools for eating symptomatology, such as the Eating Disorder Examination-Questionnaire [137].

Conclusions

Reliability generalization is a valuable method of educating other researchers about reliability issues and emphasizing that reliability is “not an immutable -unchanging property of tests” ([18], p. 124). This study indicates that test score reliability for the EDI and EAT is greater for adult and clinical samples than for adolescent and non-clinical 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 be reliably measurable 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. However, the differences between clinical eating disorder and nonclinical samples were generally larger. Without reliable measurement of these concepts

Table 3 Mean test-retest reliability estimates for the EDI and EAT

Study characteristics	# of data points K	Total sample size N	Reliability		
			Sample-weighted mean reliability (SD)	95nnnnn 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%
EAT-26	4	920	.87 (.02)	.85 - .88	100.00%

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

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.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DG developed the idea for the study, assisted with the research conceptualization and design; statistical analyses, writing and revising of the manuscript. CP obtained and coded articles for reliability generalization analyses, conducted analyses and drafted the manuscript from her doctoral dissertation under the supervision of DG and LM. SA assisted with the research conceptualization, literature research, and revising the manuscript. LM assisted with the design of the study and supervised article coding and statistical analyses. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to thank Ms. Lea Simms for her editorial assistance.

Author details

¹School of Psychology, Social Work and Social Policy, University of South Australia, Magill Campus, GPO Box 2471, Adelaide, SA 5001, Australia. ²Texas A&M University, College Station, TX, USA. ³Dickinson College, Carlisle, PA, USA.

Received: 19 November 2013 Accepted: 25 February 2014

Published: 10 March 2014

References

1. Hoek HW, Van Hoeken D: **Review of the prevalence and incidence of eating disorders.** *Int J Eat Disorder* 2003, **34**:383–396.
2. Berg KC, Frazier P, Sherr L: **Change in eating disorder attitudes and behavior in college women: prevalence and predictors.** *Eat Behav* 2009, **10**(3):137–142.
3. Touchette E, Henegar A, Godart NT, Pryor L, Falissard B, Tremblay RE, Côté SM: **Subclinical eating disorders and their comorbidity with mood and anxiety disorders in adolescent girls.** *Psychiat Res* 2011, **185**(1–2):185–192.
4. Micali N, Hagberg KW, Petersen I, Treasure JL: **The incidence of eating disorders in the UK in 2000–2009; findings from the General Practice Research Database.** *BMJ Open* 2013, **3**:e002646.
5. Hay PJ, Mond J, Buttner P, Darby A: **Eating disorder behaviors are increasing: findings from two sequential community surveys in South Australia.** *PLoS One* 2008, **3**(2):e1541.
6. Yin P, Fan X: **Assessing the reliability of Beck Depression Inventory scores: reliability generalization across studies.** *Educ Psychol Meas* 2000, **60**:201–223.
7. Thompson B: **Guidelines for authors reporting score reliability estimates.** *Educ Psycho Meas* 1994, **54**:837–847.
8. Wilkinson L, APA Task Force on Statistical Inference: **Statistical methods in psychology journals: guidelines and explanations.** *Am Psychol* 1999, **54**:594–604.
9. Thompson B: **Understanding reliability and coefficient alpha, really.** In *Score reliability: Contemporary thinking on reliability issues.* Edited by Thompson B. Thousand Oaks, CA: Sage Publications; 2003.
10. Baugh F: **Correcting effect sizes for score reliability: a reminder that measurement and substantive issues are linked inextricably.** *Educ Psychol Meas* 2002, **62**:254–263.
11. Thompson B: **Significance, effect sizes, stepwise methods, and other issues: strong arguments move the field.** *J Exp Educ* 2001, **70**:80–93.
12. Vacha-Haase T: **Reliability generalization: exploring variance in measurement error affecting score reliability across studies.** *Educ Psychol Meas* 1998, **58**:6–20.
13. Vacha-Haase T, Henson RK, Caruso JC: **Reliability generalization: moving toward improved understanding and use of score reliability.** *Educ Psychol Meas* 2002, **62**:562–569.
14. Lipsey MW, Wilson DB: *Practical meta-analysis.* Thousand Oaks: Sage Publications; 2001.
15. Hunter JE, Schmidt FL: **Correcting for sources of artificial variation across studies.** In *The Handbook of Research Synthesis.* Edited by Cooper H, Hedges LV. New York: Russell Sage Foundation; 1994:323–336.
16. Meier ST, Davis SR: **Trends in reporting psychometric properties of scales in counseling psychology research.** *J Couns Psychol* 1990, **37**:113–115.
17. Caruso JC: **Reliability generalization of the NEO Personality Scales.** *Educ Psychol Meas* 2000, **60**:236–254.
18. Henson RK, Thompson B: **Characterizing measurement error in scores across studies: Some recommendations for conducting "reliability generalization" studies.** *Meas Eval Couns Dev* 2002, **35**:113–126.
19. Breidbord J, Croudace TJ: **Reliability generalization for childhood autism rating scale.** *J Autism Dev Disord* 2013, **43**:2855–2865.
20. Miller CS, Woodson J, Howell RT, Shields AL: **Assessing the reliability of scores produced by the substance abuse subtle screening inventory.** *Subst Use & Misuse* 2009, **44**:1090–1100.
21. Vassar M, Bradley G: **A reliability generalization meta-analysis of coefficient alpha for the Reynolds Adolescent Depression Scale.** *Clin Child Psychol Psychiat* 2012, **17**:519–527.
22. Meca JS, Lopez-Pina JA, Lopez-Lopez JA, Marin-Martinez F, Rosa-Alcazar AI, Gomez-Conesa AA: **The Maudsley obsessive-compulsive inventory: a reliability generalization meta-analysis.** *Int J Clin Health Psychol* 2011, **11**:473–493.
23. Rouse SV: **Using reliability generalization methods to explore measurement error: an illustration using the MMPI-2 PSY-5 scales.** *J Pers Assess* 2007, **88**:264–275.
24. Hunter JE, Schmidt FL: *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park: Sage Publications; 1990.
25. Garner DM, Olmstead MP, Polivy J: **Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia.** *Int J Eat Disorder* 1983, **2**:15–34.
26. Garner DM, Garfinkel PE: **The Eating Attitudes Test: An index of the symptoms of anorexia nervosa.** *Psychol Med* 1979, **9**:273–279.
27. Garner DM: *Eating Disorder Inventory-2 Professional Manual.* Odessa: Psychological Assessment Resources, Inc; 1991.
28. Garner DM, Olmstead MP, Bohr Y, Garfinkel PE: **The eating attitudes test: psychometric features and clinical correlates.** *Psychol Med* 1982, **12**:871–878.
29. Gross J, Rosen JC, Leitenberg H, Willmuth ME: **Validity of the eating attitudes test and the eating disorders inventory in bulimia nervosa.** *J Consult Clin Psych* 1986, **54**:875–876.
30. Scheinberg Z, Koslowsky M, Bleich A, Mark M, Apter A, Danon Y, Solomon Z, Babur I: **Sensitivity, specificity, and positive predictive value as measures of prediction accuracy: the case of the EAT-26.** *Educ Psychol Meas* 1993, **53**(5):831–839.
31. Williamson DA, Anderson DA, Gleaves DH: **Anorexia nervosa and bulimia nervosa: structured interview methodologies and psychological assessment.** In *Body image, eating disorders, and obesity: An integrative guide for assessment and treatment.* Edited by Thompson JK. Washington, DC: American Psychological Association; 1996:205–223.
32. Cronbach LJ: **Coefficient alpha and the internal structure of tests.** *Psychometrika* 1951, **16**:197–334.
33. Dimitrov DM: **Reliability: arguments for multiple perspectives and potential problems with generalization across studies.** *Educ Psychol Meas* 2002, **62**:783–801.
34. Arthur W Jr, Bennett W Jr, Huffcutt AI: *Conducting meta-analysis using SAS.* Mahwah: Lawrence Erlbaum Associates; 2001.
35. Adkins EC, Keel PK: **Does "excessive" or "compulsive" best describe exercise as a symptom of bulimia nervosa?** *Int J Eat Disorder* 2005, **38**:24–29.
36. Birkeland R, Thompson JK, Phares V: **Adolescent motherhood and postpartum depression.** *J Clin Child Adol Psychol* 2005, **34**:292–300.
37. Compian L, Gowen LK, Hayward C: **Peripubertal girls' romantic and platonic involvement with boys: associations with body image and depressive symptoms.** *J Res Adolescence* 2004, **14**:23–47.

38. Davies K, Wardle J: **Body image and dieting in pregnancy.** *J Psychosom Res* 1994, **38**:787–799.
39. Eberenz KP, Gleaves DH: **An examination of the internal consistency and factor structure of the eating disorder inventory-2 in a clinical sample.** *Int J Eat Disorder* 1994, **16**:371–379.
40. Espelage DL, Mazzeo SE, Aggen SH, Quittner AL, Sherman R, Thompson R: **Examining the construct validity of the eating disorder inventory.** *Psychol Assess* 2003, **15**:71–80.
41. Fitzgibbon ML, Sánchez-Johnsen LA, Martinovich Z: **A test of the continuity perspective across bulimic and binge eating pathology.** *Int J Eat Disorder* 2003, **34**:83–97.
42. Franko DL, Striegel-Moore RH, Barton BA, Schumann BC, Garner DM, Daniels SR, Schreiber GB, Crawford PB: **Measuring eating concerns in black and white adolescent girls.** *Int J Eat Disorder* 2004, **35**:179–189.
43. Janelle CM, Hausenblas HA, Fallon EA, Gardner RE: **A visual search examination of attentional biases among individuals with high and low drive for thinness.** *Eat Weight Disord* 2003, **8**:138–144.
44. Joiner TE Jr, Heatherton TF, Keel PK: **Ten-year stability and predictive validity of five bulimia-related indicators.** *Am J Psychiat* 1997, **154**:1133–1138.
45. Keery H, Boutelle K, van den Berg P, Thompson JK: **The impact of appearance-related teasing by family members.** *J Adolescent Health* 2005, **2005**(37):120–127.
46. Leung F, Wang J, Tang CW: **Psychometric properties and normative data of the eating disorder inventory among 12 to 18 year old Chinese girls in Hong Kong.** *J Psychosom Res* 2004, **57**:59–66.
47. Martin KA, Hausenblas HA: **Psychological commitment to exercise and eating disorder symptomatology among female aerobic instructors.** *Sport Psycho* 1998, **12**:180–190.
48. Pelletier LG, Dion S, Lévesque C: **Can self-determination help protect women against Sociocultural influences about body image and reduce their risk of experiencing bulimic symptoms?** *J Soc Clin Psychol* 2004, **23**:61–88.
49. Podar I, Hannus A, Allik J: **Personality and affectivity characteristics associated with eating disorders: a comparison of eating disordered, weight-preoccupied, and normal samples.** *J Pers Assess* 1991, **73**:133–147.
50. Rhea DJ: **Eating disorder behaviors of ethnically diverse urban female adolescent athletes and non-athletes.** *J Adolescence* 1999, **22**:379–388.
51. Robinson TN, Killen JD, Litt IF, Hammer LD, Wilson DM, Hayward KF, Hayward C, Taylor CB: **Ethnicity and body dissatisfaction: are Hispanic and Asian girls at increased risk for eating disorders?** *J Adolescent Health* 1996, **19**:384–393.
52. Ryu HR, Lyle RM, Galer-Unti RA, Black DR: **Cross-cultural assessment of eating disorders: psychometric characteristics of a Korean version of the Eating Disorder Inventory-2 and the Bulimia Test-Revised.** *Eat Disord* 1999, **7**:109–122.
53. Shore RA, Porter JE: **Normative and reliability data for 11 to 18 year olds on the Eating Disorder Inventory.** *Int J Eat Disorder* 1990, **9**:201–207.
54. Schoemaker C, van Strien T, van der Staak C: **Validation of the eat disord inventory in a nonclinical population using transformed and untransformed responses.** *Int J Eat Disorder* 1994, **15**:387–393.
55. Tasca GA, Illing V, Lybanon-Daigle V, Bissada H, Balfour L: **Psychometric properties of the Eating Disorders Inventory-2 among women seeking treatment for binge eating disorder.** *Assessment* 2003, **10**:228–236.
56. Thurffjell B, Edlund B, Arinell H, Hägglöf B, Engström I: **Psychometric properties of Eating Disorder Inventory for children (EDI-C) in Swedish girls with and without a known eating disorder.** *Eat Weight Disord* 2003, **8**:296–303.
57. Tiggemann M: **Television and adolescent body image: the role of program content and viewing motivation.** *J Soc Clin Psychol* 2005, **24**:361–381.
58. Tylka TL, Subich LM: **A preliminary investigation of the eating disorder continuum with men.** *J Couns Psychol* 2002, **49**:273–279.
59. van Strien T, Ouwens M: **Validation of the Dutch EDI-2 in one clinical and two nonclinical populations.** *Eur J Psychol Assess* 2003, **19**:66–84.
60. Vincent MA, McCabe MP, Ricciardelli LA: **Factorial validity of the Bulimia Test-Revised in adolescent boys and girls.** *Behav Res Therapy* 1999, **37**:1129–1140.
61. Vohs KD, Bardone AM, Joiner TE Jr, Abramson LY, Heatherton TF: **Perfectionism, perceived weight status, and self-esteem interact to predict bulimic symptoms: a model of bulimic symptom development.** *J Abnormal Psychol* 1999, **108**:695–700.
62. Wassenaar D, le Grange D, Winship J, Lachenicht L: **The prevalence of eating disorder pathology in a cross-ethnic population of female students in South Africa.** *Eur Eat Disord Rev* 2000, **8**:225–236.
63. Wear RW, Pratz O: **Test-retest reliability for the eating disorder inventory.** *Int J Eat Disorder* 1987, **6**:767–769.
64. Welch G, Hall A: **The reliability and discriminant validity of three potential measures of bulimic behaviours.** *J Psychiat Res* 1987, **23**:125–133.
65. Wilhelmsson M, Andersson AL: **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).** *Eat Weight Disord* 2005, **10**:175–186.
66. Wonderlich AL, Ackard DM, Henderson JB: **Childhood beauty pageant contestants: associations with adult disordered eating and mental health.** *Eat Disord* 2005, **13**:291–301.
67. Zabinski MF, Calfas KJ, Gehrman CA, Wilfley DA, Sallis JF: **Effects of a physical activity intervention on body image in university seniors: Project GRAD.** *Ann Behav Med* 2001, **23**:247–252.
68. Breen HB, Espelage DL: **Nutrition expertise in eating disorders.** *Eat Weight Disord* 2004, **9**:120–125.
69. Brookings JB, Wilson JF: **Personality and family-environment predictors of self-reported eating attitudes and behaviors.** *J Pers Assess* 1994, **63**:313–326.
70. Botta RA: **For your health? The relationship between magazine reading and adolescents' body image and eating disturbances.** *Sex Roles* 2003, **48**:389–399.
71. Joiner GW, Kashubeck S: **Acculturation, body image, self-esteem, and eating-disorder symptomatology in adolescent Mexican American women.** *Psychol Women Quart* 1996, **20**:419–435.
72. Tylka TL, Subich LM: **Revisiting the latent structure of eating disorders: Taxometric analyses with nonbehavioral indicators.** *J Couns Psychol* 2003, **50**:276–286.
73. van den Berg P, Thompson JK, Obremski-Brandon K, Coovert M: **The Tripartite Influence model of body image and eating disturbance. A covariance structure modeling investigation testing the mediational role of appearance comparison.** *J Psychosom Res* 2002, **53**:1007–1020.
74. Vanderheyden DA, Fekken GC, Boland FJ: **Critical variables associated with bingeing and bulimia in a university population: A factor analytic study.** *Int J Eat Disorder* 1988, **7**:321–329.
75. Williams TL, Gleaves DH: **Childhood sexual abuse, body image, and disordered eating: a structural modeling analysis.** *J Trauma Dissociation* 2003, **4**:91–108.
76. Tachikawa H, Yamaguchi N, Haiyana K, Kobayashi J, Sato S, Mizukami K, Asada T, Sugie M: **The Eating Disorder Inventory-2 in Japanese clinical and non-clinical samples: Psychometric properties and cross-cultural implications.** *Eat Weight Disord* 2004, **9**:107–113.
77. Tylka TL: **The relation between body dissatisfaction and eating disorder symptomatology: an analysis of moderating variables.** *J Couns Psychol* 2004, **51**:178–191.
78. Nobakht M, Kezhkam M: **An epidemiological study of eating disorders in Iran.** *Int J Eat Disorder* 2000, **28**:265–271.
79. Sherry SB, Hewitt PL, Besser A, McGee BJ, Flett GL: **Self-oriented and socially prescribed perfectionism in the Eating Disorder Inventory Perfectionism subscale.** *Int J Eat Disorder* 2004, **35**:69–79.
80. Sim L, Zeman J: **Emotion awareness and identification skills in adolescent girls with bulimia nervosa.** *J Clin Child Adolescent Psychol* 2004, **33**:760–771.
81. Hund AR, Espelage DL: **Childhood sexual abuse, disordered eating, alexithymia, and general distress: a mediation model.** *J Couns Psychol* 2005, **52**:559–573.
82. Harrison K: **The body electric: thin-ideal media and eating disorders in adolescents.** *J Comm* 2000, **50**:119–143.
83. Siervo M, Boschi V, Papa A, Bellini O, Falconi C: **Application of the SCOFF, Eating Attitude Test 26 (EAT 26) and Eating Inventory (TFEQ) questionnaires in young women seeking diet-therapy.** *Eat Weight Disord* 2005, **10**:76–82.
84. Raciti MC, Norcross JC: **The EAT and EDI: screening, interrelationships, and psychometrics.** *Int J Eat Disorder* 1987, **6**:579–586.

85. Rizvi SL, Stice E, Agras WS: Natural history of disordered eating attitudes and behaviors over a 6-year period. *Int J Eat Disorder* 1999, **26**:406–413.
86. Piran N, Cormier HC: The social construction of women and disordered eating patterns. *J Couns Psychol* 2005, **52**:549–558.
87. Spillane NS, Boerner LM, Anderson KG, Smith GT: Comparability of the Eating Disorder Inventory-2 between women and men. *Assessment* 2004, **11**:85–93.
88. Aarnio K, Lindeman M: Magical food and health beliefs: A portrait of believers and functions of the beliefs. *Appetite* 2004, **43**:65–74.
89. Banasiak SJ, Wertheim EH, Koemer J, Voudouris NJ: Test-retest reliability and internal consistency of a variety of measures of dietary restraint and body concerns in a sample of adolescent girls. *Int J Eat Disorder* 2001, **29**:85–89.
90. Baş M, Açı H, Karabudak E, Kızıltan G: Eating attitudes and their psychological correlates among Turkish adolescents. *Adolescence* 2004, **39**:593–599.
91. Boerner LM, Spillane NS, Anderson KG, Smith GT: Similarities and differences between women and men on eating disorder risk factors and symptom measures. *Eat Behav* 2004, **5**:209–222.
92. Bittinger JN, Smith JE: Mediating and moderating effects of stress perception and situation type on coping responses in women with disordered eating. *Eat Behav* 2003, **4**:89–106.
93. Cash TF, Hrabosky JL: The effects of psychoeducation and self-monitoring in a cognitive-behavioral program for body-image improvement. *Eat Disord* 2003, **11**:255–270.
94. Cash TF, Melnyk SE, Hrabosky JL: The assessment of body image investment: an extensive revision of the appearance schemas inventory. *Int J Eat Disorder* 2004, **35**:305–316.
95. Daubennier JJ: The relationship of yoga, body awareness, and body responsiveness to self-objectification and disordered eating. *Psychol Women Quart* 2005, **29**:207–219.
96. Espina Eizaguirre A, de Cabezon AOS, de Alda IO, Olariaga LJ, Juaniz M: Alexithymia and its relationships with anxiety and depression in eating disorders. *Pers Individ Diff* 2004, **36**:321–331.
97. Favaro A, Rodella FC, Santonastaso P: Binge eating and eating attitudes among Nazi concentration camp survivors. *Psychol Med* 2000, **30**:463–466.
98. Davison KK, Markey CN, Birch LL: A longitudinal examination of patterns in girls' weight concerns and body dissatisfaction from ages 5 to 9 years. *Int J Eat Disorder* 2003, **33**:320–332.
99. Fung MSC, Yuen M: Body image and eating attitudes among adolescent Chinese girls in Hong Kong. *Percept Motor Skill* 2003, **96**:57–66.
100. Furnham A, Badmin N, Sneade I: Body image dissatisfaction: Gender differences in eating attitudes, self-esteem, and reasons for exercise. *J Psychol* 2002, **136**:581–596.
101. Graber JA, Tyrka AR, Brooks-Gunn J: How similar are correlates of different subclinical eating problems and bulimia nervosa? *J Child Psychol Psychiatr* 2003, **44**:262–273.
102. Haase AM, Prapavessis H, Owens RG: Perfectionism, social physique anxiety and disordered eating: A comparison of male and female elite athletes. *Psychol Sport Exercise* 2002, **3**:209–222.
103. Holt MK, Espelage DL: Problem-solving skills and relationship attributes among women with eating disorders. *J Counsel Deve* 2002, **80**:346–354.
104. Francis LA, Birch LL: Maternal influences on daughters' restrained eating behavior. *Health Psychol* 2005, **24**:548–554.
105. Humphrey TA, Ricciardelli LA: The development of eating pathology in Chinese-Australian women: Acculturation versus culture clash. *Int J Eat Disorder* 2004, **35**:579–588.
106. Iyer DS, Haslam N: Body image and eating disturbance among south Asian-American women: the role of racial teasing. *Int J Eat Disorder* 2003, **34**:142–147.
107. Jackson T, Weiss KE, Lunquist JJ, Soderlind A: Sociotropy and perceptions of interpersonal relationships as predictors of eating disturbances among college women: two prospective studies. *J Genet Psychol* 2005, **166**:346–359.
108. Johnson CE, Petrie TA: Relationship of gender discrepancy to psychological correlates of disordered eating in female undergraduates. *J Couns Psychol* 1996, **43**:473–479.
109. Johnson CS, Bedford J: Eating attitudes across age and gender groups: a Canadian study. *Eat Weight Disord* 2004, **9**:16–23.
110. Jordan PJ, Redding CA, Troop NA, Treasure J, Serpell L: Developing a stage of change measure for assessing recovery from anorexia nervosa. *Eat Behav* 2003, **3**:365–385.
111. Kirk G, Singh K, Getz H: Risk of eating disorders among female college athletes and nonathletes. *J Coll Counsel* 2001, **4**:122–132.
112. Koslowsky M, Scheinberg Z, Bleich A, Mark M, Apter A, Danon Y, Solomon Z: The factor structure and criterion validity of the short form of the Eating Attitudes Test. *J Pers Assess* 1992, **58**:27–35.
113. Lee S, Kwok K, Liao C, Leung T: Screening Chinese patients with eating disorders using the Eating Attitudes Test in Hong Kong. *Int J Eat Disorder* 2002, **32**:91–97.
114. Lee S, Lee AM: Disordered eating in three communities of China: a comparative study of female high school students in Hong Kong, Shenzhen, and rural Hunan. *Int J Eat Disorder* 2002, **27**:317–327.
115. Lorenzo CR, Lavori PW, Lock JD: Eating attitudes in high school students in the Philippines: A preliminary study. *Eat Weight Disord* 2002, **7**:202–209.
116. Mazzeo SE: Modification of an existing measure of body image preoccupation and its relationship to disordered eating in female college students. *J Couns Psychol* 1999, **46**:42–50.
117. McVey GL, Davis R, Tweed S, Shaw BF: Evaluation of a school-based program designed to improve body image satisfaction, global self-esteem, and eating attitudes and behaviors: A replication study. *Int J Eat Disorder* 2004, **36**:1–11.
118. McVey GL, Pepler D, Davis R, Flett GL, Abdollell M: Risk and protective factors associated with disordered eating during early adolescence. *J Early Adolescence* 2002, **22**:75–95.
119. Miotto P, De Coppi M, Frezza M, Preti A: The spectrum of eating disorders: Prevalence in an area of Northeast Italy. *Psychiat Res* 2003, **119**:145–154.
120. Moradi B, Dirks D, Matteson AV: Roles of sexual objectification experiences and internalization of standards of beauty in eating disorder symptomatology: A test and extension of objectification theory. *J Couns Psychol* 2005, **52**:420–428.
121. Ohring R, Graber JA, Brooks-Gunn J: Girls' recurrent and concurrent body dissatisfaction: Correlates and consequences over 8 years. *Int J Eat Disorder* 2002, **31**:404–415.
122. Prouty AM, Protinsky HO, Canady D: College women: eating behaviors and help-seeking preferences. *Adolescence* 2002, **37**:353–363.
123. Russell CJ, Keel PK: Homosexuality as a specific risk factor for eating disorders in men. *Int J Eat Disorder* 2002, **31**:300–306.
124. Santonastaso P, Mondini S, Favaro A: Are fashion models a group at risk for eating disorders and substance abuse? *Psychother Psychosom* 2002, **71**:168–172.
125. Scheinberg Z, Koslowsky M, Bleich A, Mark M, Apter A, Danon Y, Solomon Z, Babur I: Sensitivity, specificity, and positive predictive value as measures of prediction accuracy: The case of the EAT-26. *Educ Psychol Meas* 1993, **53**:831–839.
126. Slater A, Tiggemann M: A test of objectification theory in adolescent girls. *Sex Roles* 2002, **46**:343–349.
127. Smith MC, Thelen MH: Development and validation of a test for bulimia. *J Consult Clin Psych* 1984, **52**:863–872.
128. Tchanturia K, Katzman M, Troop NA, Treasure J: An exploration of eating disorders in a Georgian sample. *Int J Soc Psychiatr* 2002, **48**:220–230.
129. Thome J, Espelage DL: Relations among exercise, coping, disordered eating, and psychological health among college students. *Eat Behav* 2004, **5**:337–351.
130. Wichstrøm L: Psychological and behavioral factors unproductive of disordered eating: A prospective study of the general adolescent population in Norway. *Int J Eat Disorder* 2000, **28**:33–42.
131. Wood A, Waller G, Miller J, Slade P: The development of Eating Attitude Test scores in adolescence. *Int J Eat Disorder* 1992, **11**:279–282.
132. Zmijewski CF, Howard MO: Exercise dependence and attitudes toward eating among young adults. *Eat Behav* 2003, **4**:181–195.
133. Wardle J, Watters R: Sociocultural influences on attitudes to weight and eating: results of a natural experiment. *Int J Eat Disorder* 2004, **35**:589–596.
134. Vander Wal JS: Eating and body image concerns among average-weight and obese African American and Hispanic girls. *Eat Behav* 2004, **5**:181–187.

135. Hopkinson RA, Lock J: Athletics, perfectionism, and disordered eating. *Eat Weight Disord* 2004, **9**:99–106.
136. Hausenblas HA, Janelle CM, Gardner RE, Focht BC: Viewing physique slides: affective responses of women at high and low drive for thinness. *J Soc Clin Psychol* 2004, **23**:45–60.
137. Fairburn CG, Bèglin SJ: Assessment of eating disorders: interview or self-report questionnaire? *Int J Eat Disord* 1994, **16**:363–370.

doi:10.1186/2050-2974-2-6

Cite this article as: Gleaves *et al.*: Measuring eating disorder attitudes and behaviors: a reliability generalization study. *Journal of Eating Disorders* 2014 **2**:6.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

