

# Development and Psychometric Validation of an Eating Disorder-Specific Health-Related Quality of Life Instrument

Scott G. Engel, PhD<sup>1\*</sup>  
 David A. Wittrock, PhD<sup>2</sup>  
 Ross D. Crosby, PhD<sup>1,3</sup>  
 Stephen A. Wonderlich, PhD<sup>1,3</sup>  
 James E. Mitchell, MD<sup>1,3</sup>  
 Ronette L. Kolotkin, PhD<sup>4,5</sup>

## ABSTRACT

**Objective:** Health-related quality of life (HRQOL) has been used increasingly as an outcome measure in clinical research. Although the generic quality of life instruments has been used in previous research, disease-specific instruments offer greater sensitivity and responsiveness to change than generic instruments. No such disease-specific instrument is currently available that applies to eating-disordered samples.

**Method:** The current article reports on the development and validation of the Eating Disorders Quality of Life (EDQOL)

instrument, a disease-specific HRQOL self-report questionnaire designed for disordered eating patients.

**Results:** The EDQOL demonstrates excellent psychometric properties.

**Conclusion:** The application of the EDQOL as an outcome measure in eating disorder research is considered. © 2005 by Wiley Periodicals, Inc.

**Keywords:** health-related quality of life; eating disorders

(*Int J Eat Disord* 2006; 39:62–71)

## Introduction

Eating disorders impact people in a number of ways across a broad range of domains. It is increasingly well documented that patients with eating disorders have serious physical (Pomeroy & Mitchell, 2002), psychological (Godart, Flament, Lecrubier, & Jaemmet, 2000; Munoz & Amado, 1986; Ross & Ivis, 1999), social (Mitchell, Hatsukami, Eckert, & Pyle, 1985), and role functioning difficulties (Casper & Troiani, 2001; Strober, Morrell, Burroughs, Salkin, & Jacobs, 1985; Strober, Salkin, Burroughs, & Morrell, 1982). These domains of functioning have been identified by Spilker and Revicki (1996) as being central to the concept of health-related quality of life (HRQOL). Although HRQOL has become an

important outcome variable in many, if not most, areas of clinical research, eating disorder researchers have used HRQOL measures infrequently.

A small body of research has begun to investigate HRQOL in people with eating symptomatology. This research has shown that individuals with eating disorders have lower HRQOL than those without eating disorders (Hay, 2003; Keilen, Treasure, Schmidt, & Treasure, 1996; Padierna, Quintana, Arostegui, Gonzalez, & Horcajo, 2000; Spitzer, Kroenke, & Linzer, 1995). Furthermore, more severe eating-disordered symptomatology is associated with greater HRQOL impairments (Padierna et al., 2000). Although HRQOL has been shown to improve in eating disorder patients after treatment, many HRQOL domains remain impaired relative to normative population values (Padierna, Quintana, Arostegui, Gonzalez, & Horcajo, et al., 2002).

Although the above mentioned research is an excellent start in investigating the relation of eating disorders and HRQOL, it is limited by the reliance on generic measures of HRQOL. Generic measures of HRQOL assess general aspects of HRQOL that are applicable to any disease state or person. One of the strengths of generic measures of HRQOL is that they allow comparisons across diverse groups of people and diseases. One of the limitations of generic measures, however, is that they may lack sensitivity to detect differences across groups and/or responsiveness to detect changes in patients'

Accepted 12 June 2005

\*Correspondence to: Scott Engel, PhD, Neuropsychiatric Research Institute, 700 First Avenue South, Fargo, ND 58107. E-mail: sengel@nrifargo.com

<sup>1</sup> Neuropsychiatric Research Institute, Fargo, North Dakota

<sup>2</sup> Department of Psychology, North Dakota State University, Fargo, North Dakota

<sup>3</sup> University of North Dakota School of Medicine and Health Sciences, Fargo, North Dakota

<sup>4</sup> Obesity and Quality of Life Consulting, Durham, North Carolina

<sup>5</sup> Department of Community and Family Medicine, Duke University Health System, Durham, North Carolina

Published online 27 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/eat.20200

© 2005 Wiley Periodicals, Inc.

HRQOL after treatment (Fayers & Machin, 2000). The lack of sensitivity and responsiveness associated with generic instruments has led to the development of disease-specific HRQOL instruments. Disease-specific measures assess the particular concerns and common conditions related to particular disease states, such as eating disorders. Because they are more sensitive than generic instruments, disease-specific instruments will more likely demonstrate significant differences due to their ability to yield larger effect sizes (Cohen, 1988). For example, a recent study comparing obese patients with and without binge eating disorder (BED) before gastric bypass surgery has demonstrated greater effect sizes for a disease-specific measure (de Zwaan et al., 2002).

There is a paucity of research that uses quality of life (QOL) or HRQOL measures in eating disorder research. Although generic HRQOL instruments have been used with some success, the field lacks an HRQOL instrument that offers the psychometric advantages of a disease-specific instrument. The purpose of the current article is to describe the development and validation of an HRQOL measure for eating disorder patients.

## Methods

### Participants

The 538 female participants in the current study were relatively young (mean age = 21.99 years,  $SD = 8.54$ , range = 18–66 years), primarily unmarried (90.4% described themselves as single), mostly Caucasian (95%), and most described themselves as college students with part-time work (51.4%) or college students without employment (37.5%).

### Measures

**Diagnostic Measures.** *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID)*. To place participants into diagnostic groups, we used an abbreviated form of the patient version of the SCID (First, Spitzer, Gibbon, & Williams, 1995). The SCID is a well-studied and frequently used semistructured interview of DSM Axis I psychiatric disorders. In addition to the SCID, a number of probes from the Eating Disorder Examination (EDE; Carter, Stewart, & Fairburn, 2001) were implemented to help differentiate between subjective binge episodes (SBE) and objective binge episodes (OBE).

Close contact was maintained between the interviewer on the project (SGE) and a group of eating disorder assessment specialists. If any questions arose during the course of an assessment, the assessment specialists were

consulted and provided feedback and suggestions regarding how to appropriately diagnose an individual. Eleven of the SCID-based interviews were audiotaped and recoded by a master's-level eating disorder assessment specialist for reliability in the categorization of participants. Agreement between the two raters was 100%, yielding a kappa coefficient of 1.00 and thus suggesting that participants were reliably placed in their respective groups.

*Eating Attitudes Test (EAT)*. An abbreviated version of the EAT, the Eating Attitudes Test-26 (EAT-26; Garner, Olmsted, Bohr, & Garfinkel, 1982), was used as an indication of symptom severity. The EAT-26 is highly correlated with the original EAT ( $r = .98$ ), a commonly accepted measure of symptom severity in eating-disordered patients. The EAT-26 also has excellent test-retest reliability (Garner et al., 1982). A cutoff score of 20 or higher was used to allocate participants into the eating disorder group (King, 1991). The alpha coefficient for the EAT-26 in the current study was .93.

**Development of Items for the Eating Disorders Quality of Life Instrument (EDQOL)**. The generation of individual questions on the EDQOL was conducted in three primary steps: domain generation, content generation, and item generation. To develop domains, six experts in eating disorder research and treatment identified areas of HRQOL that were most important and impacted eating-disordered patients. These experts were all doctoral-level clinicians and researchers with several decades of treatment experience and many hundreds of eating disorder-related publications between them. The experts were specifically asked to think about HRQOL as it applies across a broad range of eating-disordered patients (clinical and subclinical BED, bulimia nervosa [BN], and anorexia nervosa [AN]; from very mild to very high in severity). They were also asked to make use of their familiarity with the eating disorder research as well as their clinical experience. Once these broad domains were generated, they were aggregated to remove redundancies and overlapping domains. This step produced six general domain areas: physical, psychological, financial, social, work/school, and legal.

Content generation was completed next. Five of the six experts listed relevant areas of functioning under each of the broad domain areas. Next, items meant to tap each of these content areas were generated. Each of the experts then volunteered to focus on one domain area (with two of them taking two areas) in which they were considered to have particular expertise. The raters then generated items for each of the content areas under each domain. They were specifically reminded of the multidimensional nature of many of the constructs and that many or most of them may need to be assessed with several questions. (e.g., depression would likely need a number of items to measure it adequately given its multidimensional nature.) Next, all

generated items were collected and redistributed so that each expert was able to add additional items to all of the domains. This step was completed to ensure that a comprehensive set of questions was developed for each domain area. The sixth expert, an eminent psychiatrist and researcher in the field (JEM), then carefully evaluated each of the items in each domain to again ensure that all of the domains were sufficiently assessed. Finally, generic HRQOL instruments were also reviewed and searched for items that may be applicable, but had not already been included. One hundred thirteen items were generated, with response options being never, rarely, sometimes, often, and always. Responses of never were scored as 0. Responses of always were scored as 4.

As suggested by past researchers who have developed HRQOL instruments (Juniper, Guyatt, & Jaeschke, 1996), steps were taken to pilot the generated items before participants in the study completed them. Twelve college students read the items and provided feedback regarding which items were confusing or easily misinterpreted. Next, 6 patients with diagnosed eating disorders piloted the instrument to provide feedback about the content of the items and directions for the instrument. Patients explained, in their own words, what they believed each item and the directions for the instrument meant. Items or directions associated with common misinterpretations were revised.

**Collateral Measures for Convergent and Discriminant Validity.** *SF-36 Health Survey.* To demonstrate convergent and discriminant validity, the SF-36 (Ware, Snow, Kosninski, & Reese, 1993) was administered to all subjects. The SF-36 has demonstrated good psychometric properties in the past with alpha coefficients generally above .80. Test-retest coefficients have been acceptable to good (.60–.81 for a 2-week interval; Ware, 2000). Validity information has also been good for the SF-36 with moderate correlations to other well-known HRQOL instruments. Also, factor analyses have demonstrated verification of the two-dimensional structure of the instrument (Ware, 2000). Alpha coefficients for the SF-36 subscales in the current study ranged from .80 to .87.

*Nottingham Health Profile (NHP).* To demonstrate convergent validity, the Emotional Reaction and Social Isolation subscales from the NHP (McKenna, Hunt, McEwen, Backett, & Pope, 1984) were used. The NHP has demonstrated good test-retest reliability. The patterns of correlations among its subscales suggest that the NHP, like a number of generic HRQOL instruments, measures both physical and emotional components of HRQOL (Anderson, Aaronson, & Wilkin, 1993). The alpha coefficient for the NHP in the current study was .84.

*Neuroticism.* The Goldberg (1992) 10-item self-report measure of neuroticism was used to demonstrate convergent validity with the psychological domain of the

EDQOL. The instrument has good internal consistency ( $\alpha = .86$ ) and is based on the NEO Personality Inventory Revised (NEO-PI-R), which has demonstrated excellent psychometric properties (Costa & McCrae, 1997). The alpha coefficient for neuroticism in the current study was .92.

*Beck Depression Inventory (BDI).* The BDI (Beck, Rush, Shaw, & Emery, 1979) is a commonly used and well-accepted measure of depression. It has demonstrated excellent psychometric properties for a considerable time (Beck, Steer, & Garbin, 1988). The alpha coefficient for the BDI in the current study was .92.

*Social Adjustment Scale–Self-Report (SAS-SR).* Items from the Work subscale of the SAS-SR (Weissman & Bothwell, 1976) were used for convergent validity with the social domain of the EDQOL. These items were slightly revised items from the work outside home and school sections of the instrument (both found in the Work subscale). The items that applied to “work” only or “school” only were reworded to include “work or school.” The SAS-SR has demonstrated adequate internal consistency ( $\alpha = .74$ ) with a 2-week test-retest reliability of .80. Also, the SAS-SR has demonstrated both good convergent and discriminant validity (Weissman, 2000). In the current study, the alpha coefficient for the Work/School subscale of the SAS-SR was .71.

*Financial global ratings.* To assess convergent validity for the financial domain of the EDQOL, a number of global ratings of functioning in financial areas were developed. These items were primarily generated through talking with experts in relevant areas: financial counselors, business professors, criminal justice professors, and sociology professors. The alpha coefficient for the financial scale in the current study was .77.

*Grade point average (GPA).* Self-reported GPA was collected for convergent validity with the occupation/school domain of the EDQOL.

### Procedure

The methods and procedures of the current study were approved by three different institutional review boards. Participants completed the study individually. After informed consent was obtained, participants completed the brief SCID-based interview followed by the completion of the questionnaires in the protocol. Generated EDQOL items were given first, with the remaining instruments given in random order. Some of the participants ( $n = 27$ ) completed the EDQOL items 1 week after initial administration of the instrument for purposes of test-retest reliability.

Compensation for completing the study came in two forms. Introductory psychology students were awarded five extra credit points upon the completion of the study. Other participants received \$20 for completion of the questionnaires. For the 27 participants who completed

the second administration of the questionnaire, an additional \$10 was awarded (a total of \$30).

Participants entered the study through one of three possible means: (a) ongoing research studies/clinics, (b) a survey given to introductory psychology students that identified high-risk students, and (c) students in several introductory psychology classes completed the self-report instruments for extra credit. Participants who were recruited from ongoing research studies/clinics were in eating disorder research projects at the Neuropsychiatric Research Institute (Fargo, ND) and the University of Minnesota (Minneapolis, MN).

Advertisements were also placed in the Eating Disorders Institute (an eating disorder specialty clinic in Fargo, ND) and potential participants who came to the clinic were invited to contact the researchers on this project. Because they qualified for other eating disorder research studies at these locations, nearly all of these participants met criteria for the eating-disordered group in the current study.

The brief survey of the introductory psychology students simply asked if they engaged in any compensatory behaviors or other efforts to lose weight. Students who reported compensatory behaviors or diet or exercise to lose weight were interviewed with the SCID-based interview. Also, students who reported an absence of any compensatory behaviors and a lack of diet or exercise (for the non-eating-disordered group) were interviewed.

Finally, female college students in an introductory psychology course completed the EDQOL. Most of these participants were placed in the non-eating-disordered group ( $n = 327$ ). However, 46 of them were placed in the eating-disordered group due to the fact that their EAT-26 score was 20 or higher. As suggested by King (1991), this cutoff score seems reasonable for identifying those who have significant eating symptomatology.

These three means of recruitment lead to the categorization of three groups of participants: the eating-disordered, diet and exercise, and non-eating-disordered groups. The eating-disordered group met criteria for AN, BN, BED, or a subclinical variant of one of these disorders, as outlined in the 4th ed. of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 1994). The diet and exercise group reported (on the survey to introductory psychology students) using diet and exercise to lose weight, but did not meet criteria for the eating-disordered group. These participants reported a level of eating symptomatology (as measured by the EAT-26) in between the eating-disordered and non-eating-disordered groups, suggesting that they are appropriately categorized in this "in between" group. Finally, the non-eating-disordered group did not meet criteria for an eating disorder or a subclinical variant of an eating disorder and did not report dieting or exercising in an attempt to lose weight.

### **Statistical Analysis**

A combination of both classical test theory (CTT; Anastasi & Urbina, 1997) and item response theory (IRT; Verstralen, Bechger, & Maris, 2001; Weiss & Yoes, 1991) techniques was used in item selection. The statistical software used for data analysis included SPSS (SPSS, Inc., 1999), Mplus (Muthen & Muthen, 1998-2004), and PARSCALE (Muraki & Bock, 1997).

The combined use of CTT and IRT has been implemented to develop a number of recent tests (Verstralen et al., 2001). The current study made use of both CTT concepts, such as item-to-total correlation, as well as useful information from IRT such as the application of item characteristic curves (ICCs) and item information curves (IICs). ICCs provide information regarding the level of HRQOL functioning associated with each response for any one item. IICs provide the amount (the height of the graph) and the range (the width of the graph) of information that each item provides about the latent construct that it purports to measure. Finally, test information curves provide an aggregate of the IICs for a scale informing the reader of the amount of information about the latent construct across a range of HRQOL functioning. IRT analyses allow the selection of items that provide maximal information about HRQOL across a broad range of HRQOL functioning.

Data analyses were completed in a number of steps. Frequencies were run on all items to identify and remove those items with high rates of missing data or no variance. Next, corrected item-to-total correlations were run to remove items that were not correlated or were negatively correlated with the total test score. Next, point biserial correlations were calculated between group status (eating disorder diagnosis vs. no eating disorder diagnosis) and individual items to remove items that were more frequently endorsed by asymptomatic respondents. Then, item-to-HRQOL collateral instrument correlations were calculated to identify items that associated well with HRQOL measures. Next, PARSCALE was used to provide IRT-based information. Both IICs and ICCs were used to remove items that demonstrated weak psychometric properties (i.e., items that did not discriminate well across a level of HRQOL functioning or items that do not offer a broad range and/or depth of information about HRQOL). After the initial use of IRT information, an iterative series of exploratory factor analyses were run. After the initial factor analysis, decisions regarding the number of factors were made. Further item reduction was completed based on poor factor loadings ( $<.5$ ) and cross-loadings ( $>.2$ ). Next, a second factor analysis was run. Additional item reduction was completed making use of item-to-scale correlations, the alpha coefficient for each scale, and ICCs and IICs. A final exploratory factor analysis was run with the remaining items. Final decisions regarding item selection were made based on

correlations with collateral measures. Finally, a confirmatory factor analysis (CFA) was conducted to evaluate the “higher-order” relation among scales. The CFA tested a second-order model in which items were assigned to scales, and scales were considered to be part of a higher-order construct, presumably HRQOL. This model, therefore, provides a test of whether scales can be combined to create a total aggregate score. The adequacy of the models was evaluated using the Tucker-Lewis Index (TLI), the comparative fit index (CFI), and the standardized root mean residual (SRMR). Adequate model fit was based on values > .90 for the TLI and CFI, and a SRMR < .05 (Hoyle, 1995).

This research was reviewed and approved by an institutional review board.

## Results

Generally, the eating-disordered sample ( $n = 155$ ) was younger ( $p < .01$ ), less likely to be married or divorced ( $p < .01$ ), and more likely to attend college than the non-eating-disordered ( $n = 327$ ,  $p < .01$ ) or diet and exercise sample ( $n = 56$ ,  $p < .01$ ). All group comparison analyses were run using age as a covariate and without age as a covariate. Both analyses yielded virtually identical results. For the sake of simplicity, analyses presented did not use of age as a covariate.

Exploratory factor analysis with categorical indicators (i.e., ordinal data) using MPlus software resulted in a clear four-factor solution. The final extraction method chosen was principal components and the final rotation method was Promax. Results from the factor analysis are shown in Table 1. Factor loadings within a factor are presented in bold type whereas factor loadings with other factors are not.

Factor analysis was completed on the entire sample of participants. There are several reasons why this decision was made. First, there is a precedent for using participants who ranged from very severe to extremely mild on the latent construct of interest (e.g., Kolotkin, Crosby, Kosloski, & Williams, 2001). Second, with no a priori rationale for why the factor structure should differ for eating-disordered versus non-eating-disordered individuals, it seems reasonable to include both groups in the factor analysis. Third, one of the purposes of this instrument is to assess eating disorder-specific HRQOL across the spectrum of mild impairment to serious impairment. This necessitates that individuals spanning this entire range of HRQOL functioning should be included in the factor analysis. Related to this, given that individuals will be followed long-

**TABLE 1. Factor analysis**

Subscales	Factor 1	Factor 2	Factor 3	Factor 4
<b>Psychological</b>				
Feel embarrassed	.97	.01	-.14	.06
Feel worse about self	.94	.06	-.06	-.04
Want to avoid people	.94	.01	-.01	-.05
Not get better	.77	.00	.10	.1
Feel lonely	.76	.05	.07	.11
Less interest/pleasure	.70	.17	.01	.07
Not care about self	.70	.01	.14	.05
Feel odd	.73	.04	.08	.12
Avoid eating in front of others	.69	.07	.19	-.04
<b>Physical/cognitive</b>				
Comprehend materials	.01	.87	-.05	.09
Pay attention	-.01	.86	.04	-.02
Ability to concentrate	.01	.79	-.01	.13
Cold feet/hands	.05	.67	-.01	-.01
Headache	.01	.6	.07	-.02
Weakness	.16	.55	.17	-.08
<b>Financial</b>				
Cost problems	-.12	.13	.94	-.05
Difficulty paying bills	.08	.02	.73	.16
Significant financial debt	.16	-.02	.71	.17
Need to spend/credit card	.16	-.02	.70	.14
Need to borrow money	.06	.19	.70	-.02
<b>Work/school</b>				
Leave of absence	.12	.01	-.12	.9
Low grades	-.08	-.05	.15	.81
Reduce work hours	.08	.09	0	.79
Lose job	.12	.15	.13	.71
Failure in class	.07	.13	.12	.69

itudinally with the EDQOL, the instrument must be applicable to people who begin treatment quite impaired, and end treatment far less impaired. One could still argue that the factor structure of HRQOL may differ between the eating-disordered group and the non-eating-disordered group. This, however, does not appear to be the case. Factor analysis was run on only those who reported elevated levels of eating pathology (those with full clinical or subclinical eating disorder diagnoses, dieting and exercising individuals, and those who scored above 20 on the EAT-26) and the structure was essentially identical to that which was found using the entire sample.

The second-order CFA demonstrated excellent fit indices (TLI = .99, CFI = .99, SRMR = .05), suggesting that a single higher-order latent construct, presumably HRQOL, could be measured by summing the subscales of the EDQOL.

Data analysis resulted in the specification of a 25-item instrument (the EDQOL)<sup>1</sup> consisting of four subscales: Psychological (nine items), Physi-

<sup>1</sup>To obtain a complete copy and permission to use the EDQOL, correspondence should be addressed to Dr. Scott Engel, Neuropsychiatric Research Institute, 700 First Avenue South, Fargo, ND 58107. E-mail: sengel@nrfargo.com

**TABLE 2. Convergent and discriminant validity**

Subscale	Convergent Validity			Discriminant Validity		
Psychological	Neuroticism .71	BDI .73	SF-36 Mental Component .58	SAS School/Work .35	Financial global rating .38	
Physical/Cognitive	SF-36 Role Physical .41	SF-36 General Health .45	SF-36 Vitality .48	Grade point average .05	Financial global rating .28	
Financial	Financial global ratings .64			SF-36 Mental Component .27	SF-36 Physical Component .26	
Work/School	SAS School/Work .54			SF-36 Physical Component .19	Financial Global Rating .32	

Note: SAS = Social Adjustment Scale; BDI = Beck Depression Inventory.

cal/Cognitive (six items), Financial (five items), and Work/School (five items). A list of the content of all the items retained is shown in Table 2. Scale scores are calculated by averaging the items of each scale. Total score is similarly calculated by averaging the scores of all of the items of the EDQOL. Possible ranges for each subscale and the total score are 0–4. Lower scores indicate a better QOL.

### Reliability

**Internal Consistency.** Reliability coefficients (Cronbach's alpha coefficients) for individual subscales were as follows: .86 for the Physical/Cognitive subscale, .95 for the Psychological subscale, .84 for the Work/School subscale, .86 for the Financial subscale, and .94 for the overall alpha coefficient. Also, the range of item-to-scale correlations (corrected for the influence of that item) for items in each subscale was .77–.85 for the Physical/Cognitive subscale, .58–.74 for the Psychological subscale, .62–.67 for the Work/School subscale, and .64–.76 for the Financial subscale.

**Test-Retest Reliability.** Twenty-seven participants completed 1-week test-retest administrations of the EDQOL. Test-retest intraclass correlations were .97 for

the Psychological subscale, .87 for the Physical/Cognitive subscale, .90 for the Financial subscale, .14 for the Work/School subscale, and .93 for the total score.

The Work/School subscale demonstrates a low test-retest correlation. One possible reason for this is that the items of the Work/School subscale appear to be most sensitive to quite severe levels of HRQOL impairment in eating-disordered individuals. When one examines the IRT output for the Work/School items, it is clear that the items assess HRQOL most accurately in those who are very impaired by their disorder. Therefore, individuals must be very impaired in HRQOL for the items in this subscale to change markedly. This is also demonstrated by the relatively low levels of variability on this subscale between the two administrations. When you consider that the items on this subscale are most sensitive at high levels of HRQOL impairment, it is not surprising that there is less variability on this subscale than the others and that the test-retest reliability is lower than the other subscales. Despite the fact that the Work/School subscale has low test-retest reliability, the subscale does appear to be internally consistent and demonstrates reliability (Cronbach's  $\alpha = .84$ ). Includ-

**TABLE 3. Known groups validity**

	Eating-Disordered Group (N = 155)		Diet/Exercise Group (N = 56)		Non-Eating-Disordered Group (N = 327)		F	df	p	post-Hoc Analysis	Effect Size <sup>a</sup>
	M	SD	M	SD	M	SD					
Psychological	2.20	.96	1.14	.83	0.65	.64	214.80	2,535	.001	NED<D/E,ED	.45
Physical/Cognitive	1.52	.92	0.93	.74	0.68	.54	78.07	2,535	.001	NED<D/E<ED	.23
Financial	0.49	.80	0.13	.34	0.06	.18	46.26	2,534	.001	NED,D/E<ED	.15
Work/School	0.29	.62	0.09	.25	0.04	.20	23.74	2,531	.001	NED,D/E<ED	.08
Total	1.33	.67	0.69	.48	0.42	.34	196.49	2,535	.001	NED<D/E<ED	.42

Note: NED = non-eating disorder group; D/E = diet and exercise group; ED = eating disorder group.

<sup>a</sup>Partial eta squared.

ing more participants, particularly more severely impaired participants, will likely yield more acceptable test-retest reliability scores.

**Validity**

**Convergent and Discriminant Validity.** Table 3 provides convergent and discriminant validity correlations. All correlations demonstrating convergent validity are significant at  $p < .01$ . Convergent validity can be seen in the data when theoretically similar measures correlate well with a subscale. Discriminant validity can be seen in the data when the theoretically dissimilar measures to a subscale correlate relatively low with that subscale. Convergent validity correlations for all four subscales are markedly higher than the discriminant validity correlations, providing good support for the convergent and discriminant validity of the EDQOL.

**Known Groups Validity.** One means of showing the validity of the EDQOL would be to demonstrate that groups that have been shown to differ in HRQOL would also differ on the EDQOL. Discriminant validity would be demonstrated by showing that groups differ in predicted ways on the instrument. One would expect that the eating-disordered group would show more HRQOL impairment than the diet and exercise group and the non-eating-disordered sample. Table 4 provides subscale and item means and standard deviations comparing groups of participants. All four subscale scores differ significantly between groups, with eating-disordered patients showing greater impairment on all subscales.

As previously mentioned, a disease-specific HRQOL instrument should demonstrate greater sensitivity than a generic HRQOL instrument when used with the population of subjects for which it was designed. Sensitivity can be demonstrated in the current data by demonstrating that the EDQOL significantly predicts unique variance above and beyond that predicted by a generic instrument (i.e.,

the SF-36). This was done using logistic regression predicting group status (those who met a full or subclinical diagnosis or scored 20 or higher on the EAT-26 vs. the diet and exercise group). Given that the diet and exercise group reported considerable eating, weight, and shape concerns, this test of the EDQOL is a very conservative test of incremental validity. Results showed that although the SF-36 predicted group status (Nagelkerke  $R^2 = .30$ ,  $p < .001$ ), the EDQOL predicted 22% more unique variance above and beyond that predicted by the SF-36 (Nagelkerke  $R^2$  change =  $.22$ ,  $p < .001$ ).

A final means of demonstrating validity would be to replicate the findings of Padierna et al. (2000) by showing that EDQOL scores varied predictably as a function of symptom severity. To replicate the analysis done by Padierna et al. (2000), participants were categorized into three severity groups based on their scores on the EAT-26. Because Padierna et al. used the EAT-40, the exact same groups could not be replicated, but comparable minor, moderate, and severe symptom groups were generated. Participants who scored 0–19 on the EAT-26 were placed in the minor severity group ( $n = 410$ ), those who scored 20–35 were classified in the moderate severity group ( $n = 61$ ), and those with scores of 36 and above were placed in the severe group ( $n = 56$ ). Next, analyses of variance (ANOVAs) with Tukey’s hsd post-hoc tests were completed to test for differences among groups on the four EDQOL subscales. Significant differences among severity groups were found for all subscales (all at  $p < .001$ ; see Table 4 for details). Post-hoc analyses showed that all three severity groups differed from each other on the Physical/Cognitive and Work/School subscales and total score, and that the severe and moderate severity groups differed from the minor severity group on the Psychological and Financial subscales.

Related to symptom severity and the sensitivity of the EDQOL, the EDQOL should predict significantly more symptom severity variance than the SF-36

**TABLE 4. EDQOL by symptom severity**

	Minor Symptoms (N = 410)		Moderate Symptoms (N = 61)		Severe Symptoms (N = 56)		F	df	p	Post-Hoc Analysis	Effect Size <sup>a</sup>
	M	SD	M	SD	M	SD					
Psychological	0.87	.82	2.20	.80	2.22	1.18	111.08	2,524	.001	Min<Mod,Sev	.30
Physical/Cognitive	0.74	.58	1.52	.76	1.86	1.10	90.37	2,524	.001	Min<Mod<Sev	.26
Financial	0.12	.36	0.38	.69	0.49	.87	18.49	2,520	.001	Min<Mod,Sev	.07
Work/School	0.05	.26	0.24	.49	0.48	.73	36.35	2,523	.001	Min<Mod<Sev	.12
Total	0.53	.44	1.29	.54	1.46	.85	124.77	2,524	.001	Min<Mod,Sev	.32

Note: EDQOL = Eating Disorders Quality of Life; Min = minor symptoms group (Eating Attitudes Test-26 [EAT-26] = 1–19); Mod = moderate symptoms group (EAT-26 = 20–35); Sev = severe symptoms group (EAT-26 > 35).

<sup>a</sup>Partial eta squared.

(because it is disease specific rather than a generic HRQOL instrument). The data support this requirement of a disease-specific instrument. Although the SF-36 predicts a substantial amount of the variance of symptom severity ( $R^2 = 0.42$ ,  $p < .001$ ), the EDQOL accounts for additional unique variance on the EAT-26 ( $R^2$  change = 0.25,  $p < .001$ ), again supporting the incremental validity of the scale.

## Conclusion

The primary purpose of the current study was to develop and validate an HRQOL instrument for use specifically with eating-disordered patients. Researchers and clinicians alike could potentially benefit by having an instrument that is specifically made to address the key HRQOL concerns that are associated with disordered eating. The current report describes a 25-item scale with four subscales (Psychological, Physical/Cognitive, Work/School, and Financial) and a meaningful total score. The instrument may be useful as an outcome measure in clinical research, as a means of demonstrating patient improvement (or deterioration) in treatment, as a tool to facilitate communication between healthcare providers and patients, or possibly as a number of other common uses for HRQOL instruments (Pearson Assessments, 2002).

Overall, the psychometrics of the EDQOL are quite good. The EDQOL appears to have internally consistent subscales that generally demonstrate good test-retest reliability. The EDQOL also appears to be a valid measure of HRQOL for disordered eating patients. The instrument is sensitive to group differences between disordered eating and nondisordered eating groups, it differentiates groups based on symptom severity, it explains more symptom severity and group-related variance than a generic HRQOL instrument, and it demonstrates adequate convergent and discriminant validity.

As previously mentioned, the EDQOL is more sensitive than a generic HRQOL instrument (the SF-36) and, therefore, explains a larger percentage of variance than the SF-36 when predicting group status (disordered eating vs. diet and exercise) and symptom severity. The fact that the EDQOL explains more variance than a generic instrument suggests that using the EDQOL over a generic instrument to demonstrate responsiveness to change in HRQOL may be beneficial. The extent to which the EDQOL may be responsive to change is not known and is subject to empirical testing.

Recent research has suggested that some eating-disordered individuals may be more appropriate for certain types of QOL assessments than others. Mond, Hay, Rodgers, Owen, and Beaumont (2004) reported that restricting anorexic patients, for example, may not provide useful information from QOL measures that assess patients' level of happiness (e.g., the Satisfaction with Life Scale; Diener, Emmons, Larsen, & Griffin, 1985), but instead provide clinicians more useful information when completing measures with greater objectivity (e.g., the SF-36; Gonzalez-Pinto et al., 2004). Some HRQOL researchers might argue that a person's happiness is the most important outcome measure. However, restricting anorexics typically report that further weight loss will make them happier, despite the fact that clinicians and family members can clearly see that further weight loss will lead to greater impairment or even death. One has to critically examine the specific QOL measure used in patients who lack insight or are delusional. A measure that taps functional assessment (Bohn & Fairburn, 2004) or HRQOL, such as the EDQOL, may provide useful information in such an instance.

A major strength of the current study is the psychometric properties demonstrated by the EDQOL. Measures of both reliability and validity appear to be in the range of adequate to very good. Although the test-retest reliability of the Work/School subscale is the lone exception to this strength, collecting further test-retest data on the scale with more severely impaired participants will likely improve this reliability indicator considerably.

Finally, the use of IRT in the current analysis is viewed as a considerable strength of the project. To the best of our knowledge, no HRQOL instrument has made use of IRT in its development and validation. The merits of using IRT in instrument development have been discussed elsewhere (Verstralen et al., 2001), and augmenting the item selection process with IRT-based information is viewed as a considerable strength of this instrument.

## Limitations

Despite the strengths of the current study, there are limitations. First, the current sample is not very racially diverse. It was composed of approximately 95% white, 2% African American, and 1% Native American subjects. This lack of racial diversity may impair the generalization of the data used in the current study. Future research will proactively attempt to achieve much better ethnic diversity in the sample obtained.

Second, the instrument is not generalizable to men. To date, very little is known about gender dif-



ferences in males and females with eating disorders (Gonzalez-Pinto et al., 2004; Woodside et al., 2001). Because men and women appear to have different concerns and issues related to disordered eating (Andersen, Cohn, & Holbrook, 2000), the development of the EDQOL was conducted with females in mind. In fact, the experts who generated the initial pool of items were specifically asked to consider "AN, BN, and BED patients who are female and range in severity from very mild to very severe." By generating items that apply to females from the very beginning, the generalization of the final items selected to men was likely hindered.

The third limitation of the current study has to do with method variance and the fact that all of the data collected in the development of the EDQOL were attained from self-report instruments. As pointed out by Campbell and Fiske (1959), method variance can be of particular concern in instrument development and is especially important when considering convergent and discriminant validity. Although it is beyond the scope of the current project, future research might benefit greatly by the use of multiple methods (i.e., interview or behavioral observation) and by the application of Campbell and Fiske's multitrait-multimethod matrix strategies.

Finally, the current data do not offer adequate sample size of each diagnostic group to make comparisons across these groups on the EDQOL. Although making meaningful comparisons across groups must be done very carefully (Mond et al., 2004), it has been conducted with generic instruments (e.g., Padierna et al., 2000). Not surprisingly, this work with generic instruments that lack particular sensitivity to eating-disordered individuals has not yielded consistent diagnostic group difference in HRQOL. The use of a disease-specific HRQOL instrument such as the EDQOL may, indeed, demonstrate diagnostic group differences due to the fact that it is more sensitive than a generic instrument when used with eating-disordered patients.

These limitations suggest a number of future research projects implementing the EDQOL. First, a more racially diverse sample is needed to improve the generalization of the instrument. Second, although the EDQOL demonstrated greater sensitivity than a generic instrument, the current study does not allow any assessment of the responsiveness to change of the instrument. For the EDQOL to demonstrate superiority over a generic instrument (when used with eating-disordered patients), it must show that it is more responsive to change than a generic instrument. Third,

another consideration that should be investigated is the extent to which the final items apply to BED patients. Despite the fact that the experts who generated items were asked to keep all three diagnoses (AN, BN, and BED) in mind when they generated items, it appears that many of the items may apply better to AN and BN patients and less so to BED patients. Finally, as previously suggested, collecting enough new data to perform an independent CFA would provide meaningful and important information about the factor structure of the EDQOL.

### Conclusions

The EDQOL provides an option for eating disorder clinicians and researchers to assess HRQOL using a measure that is disease specific (i.e., designed to tap the issues and concerns of eating-disordered persons). The instrument may be used in clinical research to assess treatment outcome, by third-party payers to demonstrate treatment efficacy, or by psychotherapists and/or health care providers to generate discussion that may be used to inform treatment. The instrument demonstrates excellent psychometric properties and appears to be more sensitive than a generic instrument (SF-36) when used with a disordered eating sample.

The authors thank Martina de Zwaan, Tricia Myers, and Lorraine Swan-Kremier for their help with content and item generation of potential EDQOL items; Heather Simonich, Tricia Myers, Lorraine Swan-Kremier, and Christianne Lysne for their help in assessing participants appropriately; Dr. Scott Crow for help with data collection; and Wendy Troop-Gordon and Brian Meier for allowing them to recruit participants from their classes.

---

### References

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice-Hall.
- Andersen, A., Cohn, L., & Holbrook, T. (2000). *Making weight: Men's conflicts with food, weight, shape and appearance*. Carlsbad, CA: Gurze Books.
- Anderson, R.T., Aaronson, N.K., & Wilkin, D. (1993). Critical review of the international assessments of health-related quality of life. *Quality of Life Research*, 2, 369–395.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, DC: Author.
- Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.
- Beck, A.T., Steer, R.A., & Garbin, M.G. (1988). Psychometric properties of the BDI: Twenty-five years of evaluation. *Clinical Psychology Review*, 8, 77–100.
- Bohn, K., & Fairburn, C.G. (2004, October). Addressing the problem of EDNOS (with the help of the CIA). Presented at the

- Eating Disorders Research Society Annual Meeting, Amsterdam, The Netherlands.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105.
- Carter, J.C., Stewart, A.D., & Fairburn, C.G. (2001). Eating Disorder Examination Questionnaire: Norms for young adolescent girls. *Behavior Research and Therapy*, 39(5), 625–632.
- Casper, R.C., & Troiani, M. (2001). Family functioning in anorexia nervosa differs by subtype. *International Journal of Eating Disorders*, 30, 338–342.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- de Zwaan, M., Mitchell, J.E., Monson, N., Swan-Kremeier, L., Roerig, J.L., Kolotkin, R.L., & Crosby, R.D. (2002). Two measures of health-related quality of life in morbid obesity. *Obesity Research*, 10, 1143–1151.
- Diener, E., Emmons, R.A., Larsen, R.J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49, 71–75.
- Fayers, P.M., & Machin, D. (2000). *Quality of life: Assessment, analysis and interpretation*. New York: Wiley.
- First, M.B., Spitzer, R., Gibbon, M., & Williams, J.B.W. (1995). *Structured Clinical Interview for DSM-IV Axis I disorders*. Patient edition (SCID-I/P). New York: Biometrics.
- Garner, D.M., Olmsted, M.P., Bohr, Y., & Garfinkel, P.E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, 12(4), 871–878.
- Godart, N.T., Flament, M.F., Lecrubier, Y., & Jaemmet, P. (2000). Anxiety disorders in anorexia nervosa and bulimia nervosa: Comorbidity and chronology of appearance. *European Psychiatry*, 15(1), 38–45.
- Goldberg, L.R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4, 26–42.
- Gonzalez-Pinto, A., Inmaculada, F., Rodriguez, C., de Corres Blanca, F., Sonsoles, E., Fernando, R., & Purificacion, L. (2004). Purging behaviors and comorbidity as predictive factors of quality of life in anorexia nervosa. *International Journal of Eating Disorders*, 36, 445–450.
- Hay, P. (2003). Quality of life and bulimic eating behaviors: Findings from a community-based sample. *International Journal of Eating Disorders*, 44, 434–442.
- Juniper, E.F., Guyatt, G.H., & Jaeschke, R. (1996). How to develop and validate a new health-related quality of life instrument. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 75–84). Philadelphia: Lippincott-Raven.
- Keilen, M., Treasure, T., Schmidt, U., & Treasure, J. (1996). Quality of life measurements in eating disorders, angina, and transplant candidates: Are they comparable? *Journal of the Royal Society of Medicine*, 87(8), 441–444.
- King, M.B. (1991). The natural history of eating pathology in attenders to primary medical care. *International Journal of Eating Disorders*, 10, 379–387.
- Kolotkin, R.L., Crosby, R.D., Kosloski, K.D., & Williams, G.R. (2001). Development of a brief measurement to assess quality of life in obesity. *Obesity Research*, 9(2), 102–111.
- McKenna, S.P., Hunt, S.M., McEwen, J., Backett, E.M., & Pope, C. (1984). Subjective health assessments and the perceived outcome of minor surgery. *Journal of Psychosomatic Research*, 28, 105–114.
- Mitchell, J.E., Hatsukami, D., Eckert, E.D., & Pyle, R.L. (1985). Characteristics of 275 patients with bulimia. *American Journal of Psychiatry*, 143, 482–485.
- Mond, J.M., Hay, P.J., Rodgers, B., Owen, C., & Beaumont, P.J.V. (2004). Assessing quality of life in eating disorder patients. *Quality of Life Research*, 14, 171–178.
- Munoz, R.A., & Amado, H. (1986). Anorexia nervosa: An affective disorder. *New Directions in Mental Health Service*, 31, 13–19.
- Muraki, E., & Bock, R.D. (1997). *PARSCALE: IRT item analysis and test scoring for rating-scale data* [Computer software]. New York: Scientific Software International.
- Muthen, L.K., & Muthen, B.O. (1998–2004). *Mplus user's guide* (3rd ed.) [Computer software]. Los Angeles: Muthen & Muthen.
- Padierna, A., Quintana, J.M., Arostegui, I., Gonzalez, N., & Horcajo, M.J. (2000). The health-related quality of life in eating disorders. *Quality of Life Research*, 9(6), 667–674.
- Padierna, A., Quintana, J.M., Arostegui, I., Gonzalez, N., & Horcajo, M.J. (2002). Changes in health related quality of life among patients treated for eating disorders. *Quality of Life Research*, 11, 545–552.
- Pearson Assessments. (2002). *Quality of Life Inventory* [On-line]. Available: <http://assessments.ncspearson.com/assessments/tests/qoli.htm> (retrieved August 1, 2001).
- Pomeroy, C., & Mitchell, J.E. (2002). Medical complications in anorexia nervosa and bulimia nervosa. In C.G. Fairburn & K.D. Brownell (Eds.), *Eating disorders and obesity: A comprehensive handbook* (2nd ed, pp. 194–199). New York: Guilford Press.
- Ross, H.E., & Ivis, F. (1999). Binge eating and substance use among male and female adolescents. *International Journal of Eating Disorders*, 26(3), 245–260.
- Spilker, B., & Revicki, R. (1996). Quality of life. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 185–197). Philadelphia: Lippincott-Raven.
- Spitzer, R.L., Kroenke, K., & Linzer, M. (1995). Health-related quality of life in primary care patients with mental disorders. Results from the Prime-ID 1000 study. *Journal of the American Medical Association*, 274, 1511–1517.
- SPSS, Inc. (1999). *User's guide package* [Computer software]. Chicago: Prentice Hall.
- Strober, M., Morrell, W., Burroughs, J., Salkin, B., & Jacobs, C. (1985). A controlled family study of anorexia nervosa. *Journal of Psychiatric Research*, 19, 239–246.
- Strober, M., Salkin, B., Burroughs, J., & Morrell, W. (1982). Validity of the bulimia-restrictor distinction in anorexia nervosa: Parental personality characteristics and family psychiatric morbidity. *Journal of Nervous and Mental Disease*, 170, 345–351.
- Verstralen, H., Bechger, T., & Maris, G. (2001). The combined use of classical test theory and item response theory [On-line]. Available: [http://www.cito.nl/pok/poc/eind\\_fr.htm](http://www.cito.nl/pok/poc/eind_fr.htm) (retrieved April 1, 2003).
- Ware, J.E., Jr. (2000). SF-36 health survey update. *Spine*, 25(24), 3130–3139.
- Ware, J.E., Jr., Snow, K.K., Kosninski, M., & Reese, P.R. (1993). *SF-36 health survey manual and interpretation guide*. Boston: Health Assessment Lab, New England Medical Center.
- Weiss, D.J., & Yoes, M.E. (1991). Item response theory. In R.K. Habington & J. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications* (pp. 94–109). Norwell, MA: Kluwer.
- Weissman, M.M., Sholomskas, D., & John, K. (1981). The assessment of social adjustment. An update. *Archives of General Psychiatry*, 38, 1250–1258.
- Weissman, M.M., & Bothwell, S. (1976). Assessment of social adjustment by patients self-report. *Archives of General Psychiatry*, 33, 1111–1115.
- Woodside, D.B., Garfinkel, P.E., Lin, E., Goering, P., Kaplan, A.S., Goldbloom, D.S., & Kennedy, S.H. (2001). Comparisons of men with full or partial eating disorders, men without eating disorders, and women with eating disorders in the community. *American Journal of Psychiatry*, 158, 570–574.