

# Validating the Eating Disorder Inventory-3 (EDI-3): A Comparison Between 561 Female Eating Disorders Patients and 878 Females from the General Population

Loa Clausen · Jan H. Rosenvinge · Oddgeir Friberg ·  
Kristian Rokkedal

Published online: 19 October 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** The Eating Disorder Inventory (EDI) is used worldwide in research and clinical work. The 3<sup>rd</sup> version (EDI-3) has been used in recent research, yet without any independent testing of its psychometric properties. The aim of the present study was twofold: 1) to establish national norms and to compare them with the US and international norms, and 2) to examine the factor structure, the internal consistency, the sensitivity and the specificity of subscale scores. Participants were Danish adult female patients ( $N=561$ ) from a specialist treatment centre and a control group ( $N=878$ ) was women selected from the Danish Civil Registration system. Small but significant differences were found between Danish and international, as well as US norms. Overall, the factor structure was confirmed, the internal consistency of the subscales was satisfactory, the discriminative validity was good, and sensitivity and specificity were excellent. The implications from these results are discussed.

**Keywords** Eating disorder inventory-3 · Eating disorders · EDI-3 · Psychometric validity

---

L. Clausen (✉) · K. Rokkedal  
Eating Disorders Centre,  
Regional Centre of Child and Adolescent Psychiatry,  
University Hospital of Aarhus,  
Harald Selmersvej 66,  
DK 8240 Risskov, Denmark  
e-mail: loaclus@rm.dk

J. H. Rosenvinge · O. Friberg  
Department of Psychology, University of Tromsø,  
Tromsø, Norway

O. Friberg  
Psychiatric Research Centre of Northern Norway,  
University Hospital of Northern Norway,  
Tromsø, Norway

The Eating Disorder Inventory (EDI) is a self-report questionnaire widely used both in research and in clinical settings to assess the symptoms and psychological features of eating disorders. The original version of the EDI was developed in 1983 by Garner, Olmsted, and Polivy comprising three subscales measuring eating disorder symptoms, i.e., *drive for thinness* (DT), *bulimia* (B) and *body dissatisfaction* (BD), and five more general psychological features related to eating disorders, i.e., *ineffectiveness* (IN), *perfectionism* (PE), *interpersonal distrust* (ID), *interoceptive awareness* (IA) and *maturity fears* (MF). In 1991, the EDI was enlarged from 64 to 91 items to measure additional general features related to *asceticism* (AS), *impulse regulation* (IR) and *social insecurity* (SI) (Garner 1991).

The EDI-2 discriminates reliably between patients and non clinical controls, and to some degree between patient groups (Garner 1991; Lee et al. 1998; Machado et al. 2001; Nevonon et al. 2006; van Strien and Ouwens 2003). However, cross-cultural differences have been detected (e.g., Waldherr et al. 2008; Steinhausen et al. 1992; Kordy et al. 2001; Lee et al. 1998; Tachikawa et al. 2004; Podar and Allik 2009; Clausen et al. 2009), which often manifest itself as a lack of psychometric measurement invariance across cultures. Both the original and the second version (EDI-2) have been used worldwide to screen for eating disorders in the general population, to measure treatment effect and outcome, as well as in routine clinical evaluations.

The EDI-3 represents an expansion and improvement of the earlier versions of the EDI. It consists of the same 91 questions as the EDI-2, including the same three subscales of eating disorder symptoms. The reliability of these index scores collected from eating disorder patients appears excellent (Cronbach's  $\alpha=.90-.97$ ; test-retest  $r=.98$ ) (Garner 2004; Wildes et al. 2010). Based on criticism regarding the factor structure of the EDI-2 (Limbert 2004; Muro-Sans et al. 2006; Welch et al. 1988) new factor

analyses of the sum scores yielded new subscales more congruent with recent theory and research on eating disorders (Garner 2004). Thus the EDI-3 consists of the more general, though eating disorder relevant psychological trait subscales *low self-esteem* (LSE), *personal alienation* (PA), *interpersonal insecurity* (II), *interpersonal alienation* (IA), *interoceptive deficits* (ID), *emotional dysregulation* (ED), *perfectionism* (P), *asceticism* (AS) and *maturity fear* (MF). Also, three response style indicators have been added (Garner 2004). Moreover, the new version uses the six-choice format of the EDI-2, but scores were recalibrated from a 0–3 to a 0–4 format to expand the range of summative scores to improve the psychometric properties with non-clinical populations. This generally increases the variance of item scores, and possibly changes the covariance between items.

The EDI-3 revision yields adequate convergent and discriminant validity (Cumella 2006). Since 2004 many studies have used the new EDI-3 version, yet the present study is to our knowledge the first one to independently test the factor structure, the internal consistency as well as discriminative and cross-cultural validity. Furthermore, it is claimed that the factor structure of the EDI-3 captures important clinical aspects of the psychopathology of eating disorders; yet at present, no independent test of this factor structure has been conducted. A confirmatory factor analysis approach was taken in the present study to evaluate the factor structure. In the EDI-3 manual, Garner (2004) proposes to summarize the 12 primary EDI factors in two second order factors representing: 1) a general risk factor accounting for the three primary factors *drive for thinness*, *body dissatisfaction*, and *bulimia*, and 2) a general psychological disturbance factor accounting for the remaining nine primary factors. The purpose here was to examine the fit of this model. In addition, it was compared with four alternative models in order to evaluate the suitability of Garner's model. First, it was compared with a base model specifying all factors to be independent (a null-correlated model), which it obviously should outperform. Then it was compared with a simpler second-order model specifying only one general factor, and thereafter with a more complex second-order model specifying three general factors. In the latter model, the nine primary psychological disturbance factors was split up further in two disturbance factors: one for the factors *emotional dysregulation*, *perfectionism*, *ascetism* and *interoceptive deficits* (representing rigidity and inflexibility), and another for the factors *maturity fear*, *lack of self esteem*, *feelings of insecurity*, *personal alienation* and *interpersonal alienation* (representing insecurity and estrangement). Finally, it was compared with a correlation model allowing all 12 factors to covary. Generally, correlation models always outperform second-order factor models, but at the expense of being much more complex. If

the fit of the second-order model is not substantially worse than a correlation model, the former is to prefer. Finally, a random model was specified just to check the trustworthiness of the preceding factor models. All factor models should of course outperform the random model.

In many epidemiological studies cut-off scores on for instance the *drive for thinness* subscale have been used for screening purposes, but the psychometrically based rationale and general empirical support for such cut off scores may be questioned. Another important objective of the present study was to examine the sensitivity and the specificity of the EDI-3 subscales.

Using a large patient sample and a representative sample from the general population, the present study aims 1) to establish national norms for the EDI-3 and to compare them with the US and international norms provided in the EDI-3 manual, 2) to test the internal consistency of the subscales, 3) to test the primary and second-order factor structure of the EDI-3, and 4) to examine the diagnostic accuracy for each subscale by estimating the sensitivity and the specificity of cut off scores.

## Method

### Subjects and Procedure

Female patients were recruited from the eating disorder centre at the Aarhus University Hospital in Denmark. Recruited patients were mainly given outpatient treatment. The inclusion criteria were age=18 ( $M=24.8$ ,  $SD=5.7$ , range 18–54 years), and a full or partial DSM-IV diagnosis of anorexia nervosa (AN) or bulimia nervosa (BN) determined by the Eating Disorder Examination (EDE) (Fairburn and Cooper 1993). Partial AN/BN were defined as moderate to severe eating disorders with incomplete fulfilling of diagnostic criteria, i.e., not having amenorrhea, or having eating disorder symptoms with a lower frequency or shorter duration. This is similar to the DSM category "Eating Disorders Not Otherwise Specified" (EDNOS). Patients with a binge eating disorder, a BMI=30 or who missed two or more questions in at least one of the EDI subscales were excluded. All patients were examined for comorbid psychiatric and medical disorders, the results of which are reported elsewhere (Clausen 2008; Godt 2008). Comorbidity was not an exclusion criterion if the eating disorder was severe enough to stand out as the main diagnosis. Of the active sample ( $N=561$ ), 84 had AN of whom 56 with the restricting subtype and 28 with the bulimic subtype, respectively, 202 had BN, and 275 had partial AN/BN.

Non clinical controls ( $N=2000$ ) comprised women aged 18–30 years, selected from the Danish Civil Registration

system representative of the Danish female population. They were invited, by letter, to complete the Danish version of the EDI and additional questions on paper or through the internet. 935 females responded, and 57 were excluded because they missed more than one question in one of the subscales, leaving us with a final sample of 878 respondents (44%,  $N=2000$ ). Of patients included mean age was 25.8 ( $SD=3.6$ ), mean BMI was 23.3 ( $SD=4.5$ ), 16 (1.8%) had BMI<17.5 and 51 (5.8%) had BMI>30. No interviews were performed to determine formal eating disorder diagnoses and no controls were excluded because of possible eating disorders.

In agreement with the Psychological Assessment Research, Inc., the official Danish version of the EDI-3 was translated/back-translated to ensure comparability. The study was approved by the ethics committee of Region of Central Jutland and the Danish Data Protection Agency.

### Statistical Analyses

*Descriptive, Inferential and Effect Size Statistics* Statistical analyses were carried out using SPSS version 15.0. Group comparisons were performed using t-tests. Effect sizes were calculated according to Cohen (1988) using a Cohen's  $d$  of .80, .50 and .20 to indicate a strong, medium and a small effect, respectively. One-way analyses of variance (ANOVA), with Tukey's post hoc tests, were used to compare diagnostic groups. Cronbach's alpha was used to estimate the internal consistency of the subscales using values  $=.70$  as the criterion for acceptable consistency.

*Confirmatory Factor Analyses* The confirmatory factor analyses were run in LISREL v8.80. The fit of the different factor models were evaluated according to criteria from Hu and Bentler (1999) and Marsh et al. (2004), using *RMSEA* values below  $< .06$  and *CFI* values above  $> .95$  to judge a model as an acceptable approximation to real data. Chi-square tests were not used to judge absolute fit of the models, but rather to compare which factor models best reproduced the observed correlation matrix. As most models were non-nested, they were compared using the Akaike's Information Criteria (*AIC*), which is based on the chi-square index but adds a penalty for more complex models. Lower values indicate a better fitting model.

As the EDI scores were negatively and heavily skewed, especially in the non clinical control sample, item scores were normalized in PRELIS to reduce non-normality. This resulted in a less severe skew in the control sample (Mardia's multivariate kurtosis dropped from 110.9 to 97.7), but it did not affect the non-normality in the patient sample (from 70.4 to 70.3). Adding a Satorra-Bentler correction matrix to correct

the standard errors was not possible due to the large number of items. Hence, comparisons of factor models were done within rather than between groups.

*Analyses of Sensitivity and Specificity* The possibility of using EDI-3 cut off scores to determine a clinical diagnosis of eating disorder was evaluated by conducting an analysis of sensitivity and specificity. Sensitivity indicates the proportion of true positives correctly identified (a sick patient diagnosed as sick), while specificity indicates the proportion of true negatives correctly identified (a healthy person *not* receiving a diagnosis). These proportions change as the EDI cut off score is moved up or down, and may be expressed as a ROC curve describing the diagnostic discriminatory ability across the whole range of EDI cut off scores. At one point of the ROC curve, the sensitivity and the specificity are at a maximum. The Youden's index was calculated to identify this point (Youden 1950), which is the point of the curve lying farthest away from the diagonal (chance) line. As the maximum Youden's index turned out to be relatively stable over a small range of cut off scores, two cut off scores are presented: a) primarily, the cut off score tied with the maximum Youden's index, and b) the cut off score for a Youden's index being .02 points lower than the maximum value (thus having almost comparable discrimination properties, but with different sensitivity and specificity estimates). A change window of .02 was used as the Youden's index dropped down markedly if moving the cut off score one point further. The area under the curve (AUC) values indicates how well a particular EDI subscale detects an eating disorder, i.e., its discriminatory ability. AUC is reported for each subscale within each diagnostic group. A non-discriminatory test has an AUC of .5, while a perfectly discriminating test has an AUC of 1.0. A common conception is that AUC  $> .70$  is fair,  $> .80$  is good, while  $> .90$  is excellent. A non-parametric method of constructing standard errors was used. As the EDE interview (Fairburn and Cooper 1993) was used for the patient sample only, the estimates of diagnostic accuracy are inflated if left uncorrected. This was solved by assigning a particular diagnosis of eating disorder at random to individuals in the control sample according to the prevalence rate for that particular disorder. In this study we used the generally accepted prevalence rates from two-stage community studies of 0.3% for AN, 1% for BN (Hoek and van Hoeken 2003), and 2.4% for partial AN and BN (Machado et al. 2007).

Due to the high ratio of patients to controls in the present study, yielding strongly upwardly biased base rates, positive and negative predictive values are not reported. Instead, likelihood ratios (LR) are reported indicating the

chance ratio of a positive test result in diseased individuals (true positives) to that of a positive test result in non-diseased (false positives). An unbiased post-test probability of having an eating disorder, given a specific cut off score, can be calculated by multiplying the LR with prevalence odds (Akobeng 2007), or put in a Fagan’s nomogram.

**Results**

**National Norms, and Comparisons with the US and International Norms**

All EDI-3 subscales discriminated significantly ( $p < .001$ ) and strongly (Cohen’s  $d$  ranging from .71 to 2.00) between patients and non clinical controls (see Table 1).

Table 2 shows that the mean scores of the EDI-3 subscales were different between the three diagnostic groups. Differences between all three groups are found only on the B subscale. However, each diagnostic group differentiates from the others as follows. AN patients scored higher on the MF subscale than patients with BN ( $d = .33$ ) and a partial AN/BN ( $d = .34$ ). BN patients had higher scores than patients with AN and a partial AN/BN on the DT (AN:  $d = .68$ , AN/BN:  $d = .75$ ), B (AN:  $d = 1.74$ , AN/BN  $d = 1.39$ ), and BD (AN:  $d = .96$ , AN/BN:  $d = .74$ ) subscales. BN patients displayed higher scores than partial AN/BN on the LSE ( $d = .33$ ). Patients with partial AN/BN scored lower compared to AN and BN patients on the PA (AN:  $d = .22$ , BN:  $d = .23$ ), ID (AN:  $d = .33$ , BN:  $d = .38$ ) and AS (AN:  $d = .33$ , BN:  $d = .33$ ) subscales. The three diagnostic groups were comparable on the subscales II, IA, ED, and P.

Overall, Danish control norms were significantly lower than international norms (see Fig. 1) and especially US norms, but the effect sizes were small on the subscales LSE ( $d = .21$ ), II ( $d = .27$ ), IA ( $d = .47$ ), and MF ( $d = .38$ ) whereas the remaining differences were even lower ( $d < .20$ ). Differences between Danish and US norms were significant in all subscales ( $p < .001$ ) and effect sizes were large ( $d = .99$ ) for the P subscale, medium for the II ( $d = .55$ ) and IA ( $d = .54$ ) subscales, and lower ( $d > .20$ ) for the DT ( $d = .36$ ), B ( $d = .43$ ), BD ( $d = .41$ ), PA ( $d = .46$ ), ID ( $d = .29$ ), ED ( $d = .38$ ), AS ( $d = .49$ ), and MF ( $d = .48$ ) subscales.

Figure 2 illustrates that compared to international norms, Danish patients display significantly lower scores ( $p < .01$ ) on all but two of the general subscales (i.e., the ID and ED subscales). However, effect sizes were small for the LSE ( $d = .16$ ) and AS ( $d = .18$ ) subscales, small-to-moderate for the PA ( $d = .37$ ), II ( $d = .27$ ), P ( $d = .20$ ), and MF ( $d = .26$ ) subscales, and medium for the IA ( $d = .50$ ) subscale. In the manual (Garner 2004), means for the subscales DT, B, and BD are only reported for sub-groups of patients and not the total patient population, therefore not included in Fig. 2.

Compared to US norms, five of the nine general subscales were lower in the Danish sample, with a medium effect size on the P subscale ( $d = .53$ ), a small effect size ( $d = .20$ ) on the II subscale, and a less than low effect size on the subscales PA ( $d = .19$ ), IA ( $d = .16$ ), and AS ( $d = .15$ ). The ID and ED subscales yielded higher scores in the Danish sample (i.e., ID  $d = .23$ , and ED  $d = .16$ ). No significant differences were found on the LSE and MF subscales. Moreover, on the eating disorder specific subscales (i.e., DT, B, and BN) the mean scores for the subtypes restrictive (AN-R) and bulimic (AN-B) did not differ substantially from international and US

**Table 1** Subscale sum scores for patients and normal controls

	Eating disorder patients (N=561)		Normal controls (N=878)	
	M	SD	M	SD
Drive for thinness (DT)	19.29	7.23	7.24	7.07
Bulimia (B)	14.39	9.35	2.54	4.31
Body dissatisfaction (BD)	27.89	10.07	15.34	11.27
Low self-esteem (LSE)	13.21	5.52	3.97	4.64
Personal alienation (PA)	12.46	5.38	3.99	4.45
Interpersonal insecurity (II)	10.38	5.73	4.95	4.75
Interpersonal alienation (IA)	9.41	5.56	3.72	4.26
Interoceptive deficits (ID)	18.74	7.27	5.50	5.92
Emotional dysregulation (ED)	9.08	5.68	3.48	3.95
Perfectionism (P)	10.82	5.45	5.85	4.90
Asceticism (AS)	11.87	5.89	4.04	4.27
Maturity fears (MF)	10.27	6.64	6.11	4.81

**Table 2** One-way analysis of variance for differences between diagnostic groups, significance of F and Tukey’s post-hoc test

	Anorexia nervosa (N=84)		Bulimia nervosa (N=202)		Partial AN/BN (N=275)	
	Mean	SD	Mean	SD	Mean	SD
Drive for thinness (DT)	18.70	6.28	22.29**	3.99	17.28	8.50
Bulimia (B)	7.95	7.98	21.42**	5.75	11.18	8.70
Body dissatisfaction (BD)	24.36	9.32	32.47**	7.56	25.61	10.70
Low self-esteem (LSE)	13.81	5.57	14.13*	4.97	12.35	5.77
Personal alienation (PA)	13.07	5.30	13.04*	4.88	11.84	5.69
Interpersonal insecurity (II)	10.00	5.58	11.00	5.69	10.03	5.78
Interpersonal alienation (IA)	9.27	5.40	10.09	5.35	8.95	5.73
Interceptive deficits (ID)	19.89	7.57	20.07	6.78	17.41**	7.31
Emotional dysregulation (ED)	8.85	5.08	9.47	5.71	8.87	5.84
Perfectionism (P)	11.00	4.62	11.39	5.51	10.35	5.61
Asceticism (AS)	12.89	5.99	12.79	5.21	10.88**	6.18
Maturity fears (MF)	12.19**	6.89	10.03	6.28	9.87	6.75

Post hoc Tukey’s HSD, \*  $p < .05$ , \*\*  $p < .001$

norms. Danish AN-R-patients score higher on the B subscale compared to international and US norms, but the effect size was small ( $d = .23$  and  $.32$ ). For BN patients, Danish norms on the BD subscale were significantly higher than US ( $d = .30$ ) and international norms ( $d = .41$ ). Compared to international and US norms, Danish patients with partial AN/BN had lower BD ( $d = .31$  and  $.35$ ) and DT scores ( $d = .40$  and  $.51$ ).

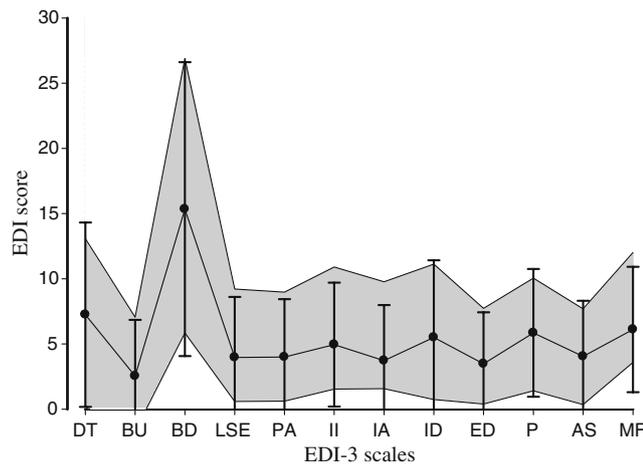
**Reliability of the EDI-3 Subscale Sum Scores**

The internal consistency of the item scores was satisfactory for patients as well as controls (see Table 3), except for the

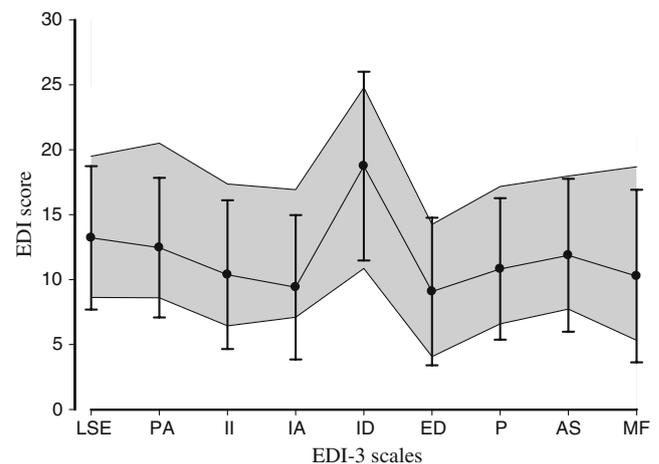
AS subscale for controls. Seven of 12 subscales for patients, and eight of 12 subscales for controls, showed an  $\alpha$  value  $> .80$ .

**Confirmatory Factor Analyses**

The confirmatory factor models were examined separately for the patient and the non clinical control sample as there is reason to expect that healthy and mentally ill individuals may attach somewhat different meanings to the same set of questions. Items were specified to load on twelve primary latent factors, according to the manual (Garner 2004). However, different ways of specifying the relationships



**Fig. 1** Norms of Danish controls vs. international controls. *Note.* Gray area displays the international norms ( $M \pm 1 SD$ ) (Garner 2004) and the error bars the Danish norms ( $M \pm 1 SD$ )



**Fig. 2** Norms of Danish patients vs. international patients. *Note.* Gray area displays the international norms ( $M \pm 1 SD$ ) (Garner 2004) and the error bars the Danish norms ( $M \pm 1 SD$ )

**Table 3** Reliability estimates (Cronbach's Alpha) of EDI-3 subscale sumscores for patients and normal controls

	Eating disorder patients (N=561)	Normal controls (N=878)
Drive for thinness (DT)	.86	.91
Bulimia (B)	.92	.87
Body dissatisfaction (BD)	.90	.93
Low self-esteem (LSE)	.86	.89
Personal alienation (PA)	.77	.83
Interpersonal insecurity (II)	.80	.83
Interpersonal alienation (IA)	.75	.79
Interoceptive deficits (ID)	.81	.85
Emotional dysregulation (ED)	.77	.78
Perfectionism (P)	.76	.80
Asceticism (AS)	.77	.59
Maturity fears (MF)	.86	.78

between the factors were tested out. The base model (M1) specified 12 independent or uncorrelated factors, which fitted the data least well as expected (see Table 4 for model comparisons). The second model specified a single general factor (M2) explaining the covariance among the 12 primary latent factors, which improved model fit according to all fit indexes. A third model (M3) defining two second order factors, nine for the psychological factors and three for the risk factors (*bulimia*, *drive for thinness* and *body dissatisfaction*), fitted the data better in terms of  $\chi^2$  and the AIC index. A tentative alternative model (M4) specifying one general risk factor, and two general psychological disturbance factors (one latent factor for *emotional dysregulation*, *perfectionism*, *asceticism* and *interoceptive deficits*, and another latent factor for the remaining disturbance factors) slightly improved model fit according to AIC. The best fitting model was, however, a 12 factor model (M5) allowing all factors to correlate freely. To test the trustworthiness of the preceding model specifications for the covariance data, a

random model (M6) specifying the 90 EDI items to load on the 12 respective factors in an unsystematic fashion produced a poorer absolute and relative fit, as expected.

Summarized, the correlated 12 factor model received best support. However, a more parsimonious second-order model, which has a much simpler factor structure than the correlation model, is to prefer if a worsening of fit is not substantial, which it was not. The difference in fit between the second-order models M2-M4 was negligible in terms of the RMSEA and the CFI. Two observations speak for favouring model M3. Firstly, the improvement in fit was larger when moving from model M2 to M3, rather than from model M3 to M4, especially in the control sample. Secondly, an examination of the factor correlations among the three general factors indicated an extremely high correlation between the two psychological factors in model M4 (.88 and .95 in the samples, respectively), while the correlations between the risk and the combined psychological factors in model M3

**Table 4** Comparison of factor models in eating disorder patients and normal controls

Model	Eating disorder patients (N=561)					Normal controls (N=878)				
	<i>df</i>	$\chi^2$	AIC	$\varepsilon_a$	CFI	<i>df</i>	$\chi^2$	AIC	$\varepsilon_a$	CFI
M1 Uncorr	3915	17986	18346	.0801	.921	3915	30355	30715	.0878	.948
M2 One 2 <sup>nd</sup>	3903	11768	12152	.0600	.942	3903	13960	14344	.0542	.970
M3 Two 2 <sup>nd</sup> (Garner's model)	3902	11441	11827	.0587	.945	3902	13462	13847	.0529	.972
M4 Three 2 <sup>nd</sup>	3900	11168	11558	.0577	.946	3900	13304	13694	.0524	.972
M5 Corr	3849	10614	11106	.0560	.948	3849	12541	13033	.0508	.974
M6 Random	3849	31808	32300	.1139	.874	3849	42592	43083	.1071	.937

*Df* = Degrees of freedom,  $\chi^2$  = Chi-square, AIC = Akaike's Information Criteria,  $\varepsilon_a$  = Root mean square error of approximation (RMSEA), CFI = Comparative Fit Index. Models: M1 = An uncorrelated 12 factor model, M2 = One second order factor, M3 = Two second order factors, as published by Garner (2004) in the manual, M4 = Three second order factors, M5 = A correlated 12 factor model, and M6 = Items specified to load on the 12 different factors in a random fashion. All factors were allowed to correlate, as in model M5.

were substantially lower (.54 and .60). Taken together eating problems should be summarized in two main scores to differentiate eating problems: one representing a risk factor and another representing a psychological disturbance score, according to the author (i.e. Garner 2004). The factor loadings from second order factor analysis of model M3 are displayed in Table 5. At the same time the fit of the preferred model (M3), as well as the best model (M5), was not great according to the *RMSEA* index. Although the *RMSEA* was lower than  $< .06$  as recommended by Hu and Bentler (1999), hence indicating a reasonably approximation of the model to the observed data, there is room for improvements. The relatively large number of chi-squares compared to degrees of freedom, is not reassuring either. Following an inspection of the modification indices, the mediocre fit appears related to several items showing hugely correlated residuals as well as significant factor side-loadings. Hence some of the EDI items do not have adequate psychometric properties.

Sensitivity and Specificity

ROC curves for all EDI-3 subscales were expressed for a diagnosis of AN, BN and partial AN/BN (see Figs. 3, 4 and 5). In each figure, the subscales with the highest AUC (Area Under Curve) value are listed first. The figures show that the *interoceptive deficits* subscale is the best predictor across all diagnostic groups, followed by *low self-esteem* and *personal alienation*. The bulimia subscale comes sixth overall, but is an excellent predictor of a diagnosis of BN with high sensitivity and specificity estimates. Table 6 provides an overview of sensitivity, specificity, likelihood ratios and diagnostic accuracy of the three best and the worst predictors within each diagnostic group. The cut off

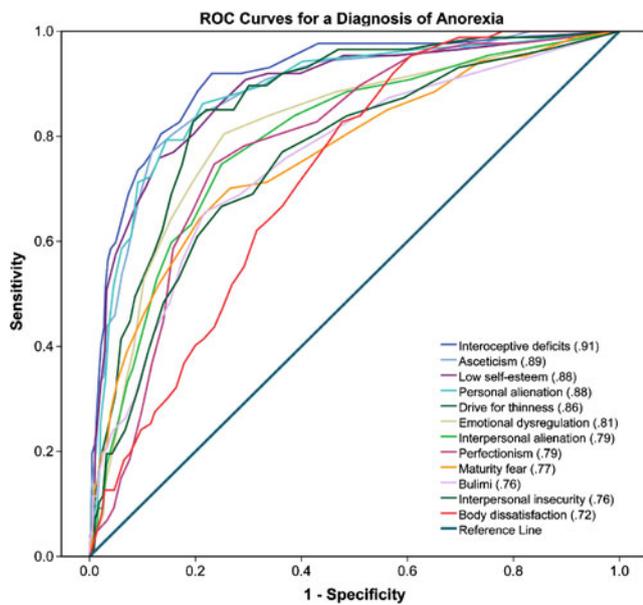
score for deciding these estimates was based on the highest value on the Youden’s index. As several of the subscales changed the Youden’s index minimally by either lowering or increasing the cut off, alternative cut off scores are also reported in the direction with the smallest change in the Youden’s index. Generally, increasing the cut off increases the specificity and reduces misclassification, but at the cost of increasing the number of false negatives (patients not detected), which represents a more serious error. Most ROC curves across the diagnostic groups are quite parallel over all levels of cut off scores, but with one notable exception. As expressed in Fig. 3, the subscale of *body dissatisfaction* is the worst of all subscales in overall diagnostic accuracy of AN. However, at low cut off scores ( $< 6$ ) it definitely is the most sensitive subscale in detecting true cases of AN, though performing poorly with regards to specificity ( $< .22$ ).

Conclusion

Overall the new version of the Eating Disorder Inventory (EDI-3) stands out as a psychodiagnostic assessment tool that may be used to capture eating problems. Apart from one subscale with a medium effect size difference, all differences between the patient and the non clinical control group yielded high effect sizes, and even slightly higher than using the EDI-2 (Clausen et al. 2009). Thus the discriminative validity is good, as is the case for internal consistency (Table 3). The latter is even better in this study than in the original development of the EDI-3 (Garner 2004). Thus, one argument for creating a new EDI-version (i.e., a more consistent measure) is supported by our findings.

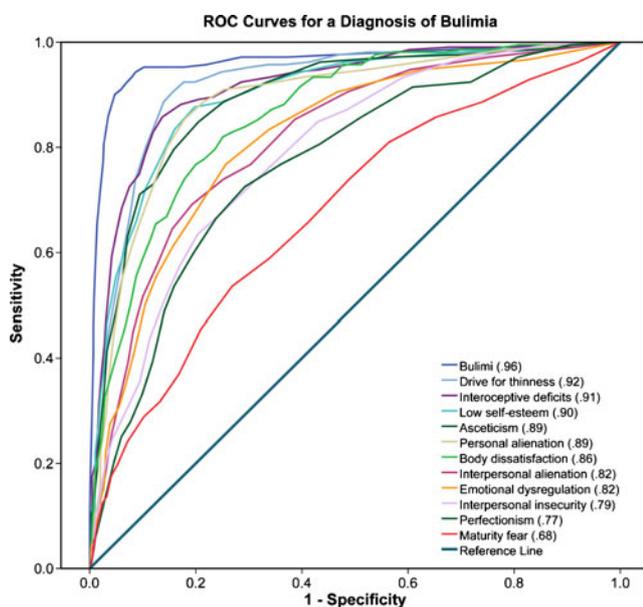
**Table 5** The factor loadings for the second order two factor model (M3 in Table 4) with risk and psychological disturbance as general factors accounting for the 12 primary factors. The two general factors were allowed to correlate

	Eating disorder patients (N=561)		Normal controls (N=878)	
	Risk	Psychological disturbance	Risk	Psychological disturbance
Drive for thinness	.92		.97	
Body dissatisfaction	.85		.83	
Bulimia	.49		.74	
Maturity fears		.51		.74
Interoceptive deficits		.83		.92
Low self-esteem		.91		.92
Perfectionism		.51		.56
Interpersonal insecurity		.60		.74
Emotional dysregulation		.67		.85
Asceticism		.82		.88
Personal alienation		1.00		.99
Interpersonal alienation		.75		.91

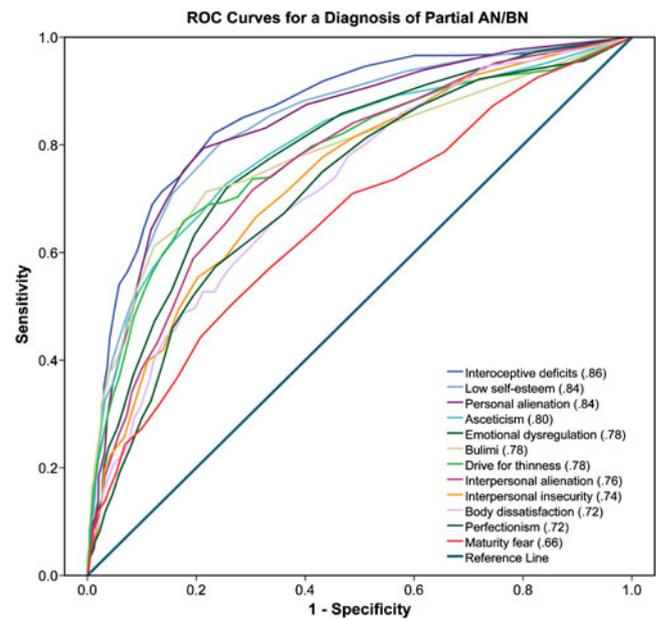


**Fig. 3** ROC Curves for a Diagnosis of Anorexia. *Note.* AUC = Percent of total area under ROC curve. A low cut-off score starts in the right upper corner, going down the diagonal

Through the present large scale study, national norms have been successfully established. For practical purposes, the implications from this study are that outside the US, the international norms (Garner 2004) may be used for screening purposes when national norms are lacking. On the other hand, the lack of national *clinical* norms may lead to a more valid comparison to US than international data. This certainly creates practical problems in doing



**Fig. 4** ROC Curves for a Diagnosis of Bulimia. *Note.* AUC = Percent of total area under ROC curve. A low cut-off score starts in the right upper corner, going down the diagonal



**Fig. 5** ROC Curves for a Diagnosis of Partial AN/BN. *Note.* AUC = Percent of total area under ROC curve. A low cut-off score starts in the right upper corner, going down the diagonal

epidemiological research, and may point to variations in how a psychological phenomenon (e.g., eating disorder problems) appear in various cultures and populations, as well as to psychometric challenges in increasing the construct validity.

The confirmatory factor analyses by and large supported the grouping of eating problems in two general factor scores, one assessing a risk component and the other assessing associated psychological disturbances. The model fit in the present study was actually better than what was presented in the EDI-3 manual (Garner 2004). One reason for this may be that Garner based his analyses on twelve subscale sum scores rather than 90 item scores, as was done in the present study. Still, the model fit was in the upper window of what is regarded as a minimal acceptable model approximation. This may be explained by the fact that several items had poor psychometric properties according to the modification indices provided by LISREL, showing hugely correlated error covariances and significant factor side-loadings. These items are thus ambiguous indicators of eating problems, and should be revised or removed in a future version of the EDI-3. Identifying these items requires, however, an extensive item level analysis followed by a cross-validation on a holdout sample. This is a task for another paper.

In our study the sensitivity and specificity estimated make the bulimia subscale an excellent predictor of a BN-diagnosis. However, compared to the EDI-2, the EDI-3 version of this subscale contains only one new item.

**Table 6** Sensitivity, specificity, likelihood rates and diagnostic accuracy of the three best and the worst EDI-3 subscales for each diagnostic group

Diagnosis/subscales	Cut-off score	AUC ( <i>CI<sub>95</sub></i> )	Sensitivity %	Specificity %	LR+	Youden's index	FN %	FP %	Misclassification %
<b>Anorexia nervosa</b>									
Interoceptive deficits	<b>9</b>   <i>12</i>	.911 (.878)	<b>.92</b>   .80	.77   .87	<b>4.0</b>   <i>6.0</i>	<b>.69</b>   .67	<b>8.0</b>   <i>19.5</i>	<b>23.1</b>   <i>13.5</i>	<b>21.7</b>   <i>14.0</i>
Ascetism	7   <b>9</b>	.886 (.849)	.84   <b>.77</b>	.80   <b>.88</b>	<i>4.1</i>   <b>6.5</b>	.64   <b>.65</b>	<i>16.1</i>   <b>23.0</b>	<i>20.3</i>   <b>11.8</b>	<i>20.0</i>   <b>12.8</b>
Low self-esteem	6   <b>10</b>	.884 (.844)	.91   <b>.76</b>	.71   <b>.87</b>	<i>3.1</i>   <b>5.8</b>	.61   <b>.63</b>	<i>9.2</i>   <b>24.1</b>	<i>29.5</i>   <b>13.1</b>	<i>27.7</i>   <b>14.1</b>
Body dissatisfaction	<i>11</i>   <b>15</b>	.722 (.677)	.95   <b>.83</b>	.39   <b>.52</b>	<i>1.6</i>   <b>1.7</b>	.35   <b>.35</b>	<i>4.6</i>   <b>17.2</b>	<i>60.7</i>   <b>47.7</b>	<i>55.6</i>   <b>44.9</b>
<b>Bulimia nervosa</b>									
Bulimia	<b>9</b>   <i>13</i>	.963 (.946)	<b>.94</b>   .90	.91   .95	<b>10.9</b>   <i>19.1</i>	<b>.86</b>   .85	<b>5.7</b>   <i>10.0</i>	<b>8.6</b>   <i>4.7</i>	<b>8.1</b>   <i>5.7</i>
Drive for thinness	<b>16</b>   <i>17</i>	.918 (.896)	<b>.91</b>   .89	.84   .86	<b>5.8</b>   <i>6.4</i>	<b>.75</b>   .75	<b>9.5</b>   <i>11.4</i>	<b>15.7</b>   <i>13.9</i>	<b>14.4</b>   <i>13.4</i>
Interoceptive deficits	<b>12</b>   <i>13</i>	.914 (.892)	<b>.86</b>   .83	.86   .88	<b>6.3</b>   <i>6.9</i>	<b>.72</b>   .71	<b>14.2</b>   <i>17.1</i>	<b>13.7</b>   <i>12.0</i>	<b>13.8</b>   <i>13.0</i>
Maturity fear	5   <b>9</b>	.679 (.638)	.81   <b>.54</b>	.43   <b>.73</b>	<i>1.4</i>   <b>2.0</b>	.25   <b>.27</b>	<i>19.0</i>   <b>46.4</b>	<i>56.5</i>   <b>26.8</b>	<i>49.2</i>   <b>30.6</b>
<b>Partial AN/BN</b>									
Interoceptive deficits	<b>9</b>   <i>13</i>	.862 (.837)	<b>.82</b>   .69	.77   .88	<b>3.5</b>   <i>5.8</i>	<b>.59</b>   .57	<b>17.9</b>   <i>31.1</i>	<b>23.2</b>   <i>11.9</i>	<b>21.9</b>   <i>16.8</i>
Low self-esteem	7   <b>9</b>	.843 (.816)	.80   <b>.71</b>	.75   .84	<b>3.3</b>   <i>4.5</i>	<b>.56</b>   .55	<b>19.6</b>   <i>29.1</i>	<b>24.5</b>   <i>15.6</i>	<b>23.2</b>   <i>19.1</i>
Personal alienation	7   <b>8</b>	.838 (.811)	.79   <b>.75</b>	.79   .82	<b>3.7</b>   <i>4.3</i>	<b>.58</b>   .58	<b>20.6</b>   <i>24.7</i>	<b>21.4</b>   <i>17.7</i>	<b>21.2</b>   <i>19.5</i>
Maturity fear	6   <b>9</b>	.657 (.620)	.71   <b>.50</b>	.51   <b>.74</b>	<i>1.5</i>   <b>1.9</b>	.22   <b>.24</b>	<i>29.1</i>   <b>49.7</b>	<i>48.7</i>   <b>26.5</b>	<i>43.6</i>   <b>32.4</b>

Cut-off scores with the highest Youden's index value in bold. Cut-off scores within a range of  $\pm .02$  on the Youden's index in italic. *LR* + Likelihood ratio, *FN* False negatives and *FP* False positive

However, the overall purpose of the EDI-3 was to compose subscales with a conceptual content more congruent with domains identified by modern thinking about the nature of eating disorders (Garner 2004). The ROC-analyses support the success of this purpose in the sense that the subscale *interoceptive deficits* is the best predictor across all diagnostic groups, followed by *low self-esteem* and *personal alienation*. Previous studies have also found interoceptive awareness along with the three eating disorder specific subscales to discriminate between eating disorder patients and psychiatric controls (Nevonen et al. 2006; Schoemaker et al. 1997). Also, interoceptive issues are related to other psychological constructs of eating disorders like depression, perfectionism, and self directiveness (Fassino et al. 2004). Hence, interoceptive deficits stand out as a concept with a high discriminative and construct validity related to eating disorders. An important implication from the present findings is that the current use of the *drive for thinness* subscale as a screening tool in epidemiological studies is clearly not warranted any more. While people scoring high on drive for thinness may do this for good as well as bad reasons even unrelated to the pathology of eating disorders, disturbance in the accuracy of perception or recognition of bodily states is an important pathognomic sign of the specific eating disorder psychopathology, commonly seen as a failure to recognize signs of hunger (Bruch 1962). Also noted by Bruch (1962) the all-pervading sense of ineffectiveness in patients with AN may be well captured by the EDI-3. This kind of ineffectiveness may reflect a personality develop-

ment attributed to a failure of confirmation of child initiated behaviour.

As the ROC-analyses indicated that the EDI-3 is highly suitable for screening purposes, a future study aimed at finding the most optimal items for screening purposes is clearly indicated as well. The present study suggests that the current version of the EDI has equally well, if not better, psychometric quality compared to what the author (i.e. Garner 2004) has reported. One caveat of performing new factor analyses on an item level to identify psychometrically poorly working items, is that removal (or revision) of items will affect the current estimates of sensitivity and specificity.

The strength of the present study is the use of a large control sample of women stratified from the general population. This stands in contrast to the common practice of using smaller student samples as controls, where the questionable representativity of the general population may deflate the external validity of the findings. Using a large population sample creates on the other hand problems in terms of case detection. Another caveat of the study is that the recording of medical complications, notably in the anorexia nervosa subsample, was incomplete, which may increase the risk of inflated scores due to the impact of malnutrition. On the other hand, this problem is relevant for less than 10% ( $N=561$ ) of the active sample.

**Acknowledgment** The authors are indebted to Michael Wynn for a thorough linguistic editing of the final version of this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Akobeng, A. K. (2007). Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatrica*, *96*, 487–491.
- Bruch, H. (1962). Perceptual and conceptual disturbances in anorexia nervosa. *Psychosomatic Medicine*, *24*, 187–194.
- Clausen, L. (2008). Time to remission for eating disorder patients: a 2½-year follow-up study of outcome and predictors. *Nordic Journal of Psychiatry*, *62*, 151–159.
- Clausen, L., Rokkedal, K., & Rosenvinge, J. H. (2009). Validating the Eating Disorder Inventory (EDI-2) in two Danish samples: a comparison between female eating disorders patients and females from the general population. *European Eating Disorders Review*, *17*, 462–467.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates, Inc.
- Cumella, E. J. (2006). Review of the Eating Disorder Inventory-3. *Journal of Personality Assessment*, *87*, 116–117.
- Fairburn, C. G., & Cooper, Z. (1993). The Eating Disorder Examination (12th ed). In C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: Nature, assessment and treatment* (pp. 317–331). NY: The Guilford Press.
- Fassino, S., Piero, A., Gramaglia, C., & Abbate-Daga, G. (2004). Clinical, psychopathological and personality correlates of interoceptive awareness in anorexia nervosa, bulimia nervosa and obesity. *Psychopathology*, *37*, 168–174.
- Garner, D. M. (1991). *Eating disorder inventory-2. Professional manual*. Odessa, FL: Psychological Assessment Research, Inc.
- Garner, D. M. (2004). *Eating Disorder Inventory-3. Professional Manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Garner, D. M., Olmsted, M. P., & Polivy, J. (1983). Development and validation of a multidimensional EDI for anorexia nervosa and bulimia. *International Journal of Eating Disorders*, *2*, 15–34.
- Godt, K. (2008). Personality disorders in 545 patients with eating disorders. *European Eating Disorders Review*, *16*, 94–99.
- Hoek, H. W., & van Hoeken, D. (2003). Review of the prevalence and incidence of eating disorders. *International Journal of Eating Disorders*, *34*, 383–396.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- Kordy, H., Percevic, R., & Martinovich, Z. (2001). Norms, normality, and clinical significant change: implications for the evaluation of treatment outcomes for eating disorders. *International Journal of Eating Disorders*, *30*, 176–186.
- Lee, S., Lee, A. M., & Leung, T. (1998). Cross-cultural validity of the Eating Disorder Inventory: a study of Chinese patients with eating disorders in Hong Kong. *International Journal of Eating Disorders*, *23*, 177–188.
- Limbert, C. (2004). The Eating Disorder Inventory: a test of the factor structure and internal consistency in a nonclinical sample. *Health Care for Women International*, *25*, 165–178.
- Machado, P. P., Goncalves, S., Martins, C., & Soares, I. C. (2001). The portuguese version of the eating disorders inventory: evaluation of its psychometric properties. *European Eating Disorders Review*, *9*, 43–52.
- Machado, P. P., Machado, B. C., Goncalves, S., & Hoek, H. W. (2007). The prevalence of eating disorders not otherwise specified. *International Journal of Eating Disorders*, *40*, 212–217.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341.
- Muro-Sans, P., Amador-Campos, J. A., & Peró-Cebollero, M. (2006). Factor structure of eating disorders inventory-2 in a Spanish sample. *Eating and Weight Disorders*, *11*, 42–52.
- Nevonen, L., Clinton, D., & Norring, C. (2006). Validating the EDI-2 in three Swedish female samples: eating disorders patients, psychiatric outpatients and normal controls. *Nordic Journal of Psychiatry*, *60*, 44–50.
- Podar, I., & Allik, J. (2009). A cross-cultural comparison of the eating disorder inventory. *International Journal of Eating Disorders*, *42*, 346–355.
- Schoemaker, C., Verbraak, M., Breteler, R., & van der Staak, C. (1997). The discriminant validity of the Eating Disorder Inventory-2. *British Journal of Clinical Psychology*, *36*, 627–629.
- Steinhausen H. C., Neumarker K. J., Vollrath M., Dudeck U., & Neumarker, U. (1992). A transcultural comparison of the eating disorder inventory in former East and West-Berlin. *International Journal of Eating Disorders*, *12*, 407–416.
- Tachikawa, H., Yamaguchi, N., Hatanaka, K., Kobayashi, J., Sato, S., Mizukami, K., et al. (2004). The Eating Disorder Inventory-2 in Japanese clinical and non-clinical samples: psychometric properties and cross-cultural implications. *Eating & Weight Disorders*, *9*, 107–113.
- van Strien T., & Ouwens, M. (2003). Validation of the Dutch EDI-2 in one clinical and two nonclinical populations. *European Journal of Psychological Assessment*, *19*, 66–84.
- Waldherr, K., Favaro, A., Santonastaso, P., van Strien, T., & Rathner, G. (2008). Comparison of the Eating Disorder Inventory (EDI) in the Netherlands, Austria and Italy. *European Eating Disorders Review*, *16*, 472–479.
- Wildes, J. E., Ringham, R. M., & Marcus, M. (2010). Emotion avoidance in patients with anorexia nervosa: Initial test of a functional model. *International Journal of Eating Disorders*, *43*, 398–404.
- Welch, G., Hall, A., & Walkey, F. (1988). The factor structure of the eating disorders inventory. *Journal of Clinical Psychology*, *44*, 51–56.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32–35.