Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/121133

Psychometric Properties of the Dutch Version of the Eating Disorder Inventory–3

Vicky Lehmann[1,2], Machteld A. Ouwens[3], Johan Braeken[2], Unna N. Danner[4], Annemarie A. van Elburg[4,5], Marrie H. J. Bekker[2], Annette Breurkens[6], and Tatjana van Strien[7]

Abstract

The psychometric properties of the Dutch version of the Eating Disorder Inventory–3 (EDI-3) were tested in eating disordered patients (*N* = 514) using confirmatory factor analyses, variance decomposition, reliabilities, and receiver operating characteristic (ROC) curve analyses. Factorial validity results supported the 12 subscales, but model fit was impaired by correlated item errors, misallocated items, and redundant subscales. At the composite level, the Bulimia subscale was identified as a largely specific source of information that did not contribute much to its overarching composite. Reliabilities for subscales and composites ranged from .6 to .9. ROC curve analysis indicated good to excellent discriminative ability of the EDI-3 identifying clinical subjects against a reference group. In conclusion, further revisions of the EDI-3 might target the item allocation and (over-)differentiation of subscales and composites to further clarify its structure. For the clinical practice, we advise the careful use of the EDI-3, although it might serve as a good screening tool.

Introduction

The Eating Disorder Inventory (EDI) is a widely used self-report measure to assess attitudes and behaviors concerning eating, weight, and shape as well as psychological traits relevant to eating disorders. Currently, the EDI is available in its third version (EDI-3; Garner, 2004) and has been developed as an alternative way to evaluate the items and outcome of the former EDI-2 by rearranging 90 of the original 91 items into partly new subscales. This reorganization was initiated due to a problematic factor structure and high correlations between subscales of the EDI-2 (Garner, 2004).

In his initial validation study, Garner (2004) investigated the psychometric properties of the EDI-3 in a clinical sample from the United States and an international clinical sample from Australia, Canada, Italy, and the Netherlands. He derived 12 subscales combining clinical expertise and exploratory factor analyses (EFA). Subsequently, he used the outcomes on the 12 subscales to cluster them to 5 composites based on second-order confirmatory factor models.

The 12 first-order subscales are as follows: Drive for Thinness (DT), Bulimia, and Body Dissatisfaction (BD), which can be summarized into one second-order composite, supposedly measuring a general risk to develop or have an eating disorder, called Eating Disorder Risk Composite (EDRC). The nine remaining first-order subscales evaluate psychological aspects relevant for eating disorders: Low Self-Esteem (LSE), Personal Alienation (PA), Interpersonal Insecurity (II), Interpersonal Alienation (IA), Interoceptive Deficits (ID), Emotional Dysregulation (ED), Perfectionism, Asceticism, and Maturity Fears (MF). They can be summarized into four more second-order composites: Ineffectiveness Composite (IC), Interpersonal Problems Composite (IPC), Affective Problems Composite (APC), and Overcontrol Composite (OC). The latter four second-order composites can be optionally combined into one General Psychological Maladjustment Composite (GPMC). A graphical

[1]University of Groningen, The Netherlands
[2]Tilburg University, The Netherlands
[3]GGz Breburg, Tilburg, The Netherlands
[4]Altrecht Eating Disorders Rintveld, Zeist, The Netherlands
[5]University Medical Center Utrecht, The Netherlands
[6]GGz Centraal, Meerkanten, The Netherlands
[7]Radboud University, Nijmegen, The Netherlands

Corresponding Author:
Vicky Lehmann, Department of Health Psychology Research, University Medical Center Groningen, FA12, A. Deusinglaan 1, 9713 AV Groningen, The Netherlands.
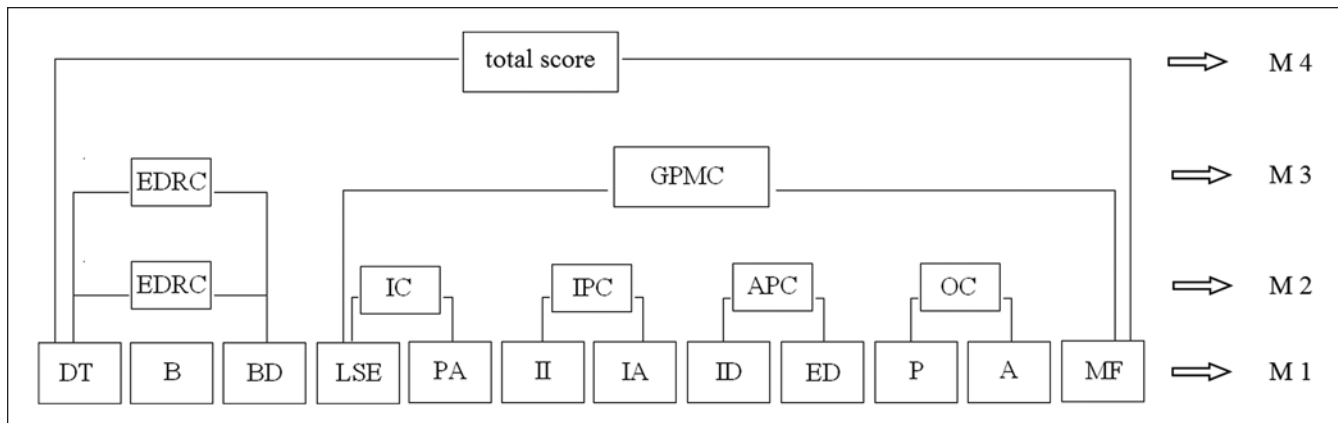Email: v.lehmann@umcg.nl; vicky.lehmann@gmail.com

Figure 1. Schematic overview of tested models in the confirmatory factor analyses.
Note. EDRC = Eating Disorder Risk Composite; GPMC = General Psychological Maladjustment Composite; IC = Ineffectiveness Composite; IPC = Inter-personal Problems Composite; APC = Affective Problems Composite; OC = Overcontrol Composite; DT = Drive for Thinness; B = Bulimia; BD = Body Dissatisfaction; LSE = Low Self-Esteem; PA = Personal Alienation; II = Interpersonal Insecurity; IA = Interpersonal Alienation; ID = Interoceptive Deficits; ED = Emotional Dysregulation; P = Perfectionism; A = Asceticism; MF = Maturity Fears.

representation of the official structure of the EDI-3 is given in Figure 1, marked as M2 and M3.

As mentioned above, Garner used an American and inter-national sample to test his factor structure, thereby demon-strating the usefulness of the EDI-3 in different languages. However, due to his two-stepped approach of first testing the item affiliation to the 12 subscales based on EFA and subsequently only conducting second-order confirmatory factor analysis (CFA) on an aggregate level (i.e., sum scores of the 12 subscales), he never comprehensively tested whether the item affiliation withstands in the second-order models. He also never tested his full model of 12 subscales and 5 composites. Instead, he presented CFAs separately for the EDRC (with 3 affiliated subscales) and the remaining psychological second-order composites.

To date, there are only three independent validation stud-ies that examined the new EDI-3 structure. Elosua and López-Jáuregui (2011) followed Garner's approach using aggregated subscale scores and found support for the internal structure of the EDI-3 (Spanish version) at the second-order composite level. It should be noted positively that they went a step further and tested eating disorder specific and psycho-logical composites in one model. However, by conducting the CFA at an aggregate level, the affiliation of the 90 items to the 12 subscales was taken for granted and no information about the factorial validity of the 12 subscales at the item level was given. Garcia-Grau and colleagues (2010) investi-gated all three versions of the EDI (EDI, EDI-2, and EDI-3; Spanish version) using a large but less generalizable sample of exclusively female secondary school students (*n* = 738). They performed their analyses at item level and could not replicate any of the factor structures originally proposed by Garner. EFA led them to put forward a structure including seven first-order factors. Clausen, Rosenvinge, Friborg, and Rokkedal (2011, Danish version) also studied the

psychometric properties of the EDI-3 at item level in a large clinical (*n* = 561) and control (*n* = 878) sample and found good results concerning reliability, specificity, and sensitiv-ity. On the second-order level, no model objectively outper-formed the simple first-order model with 12 correlated factors. Nevertheless, for reasons of parsimony, the authors argue for two second-order factors corresponding to the EDRC and the GPMC.

Next to the change in the subscale organization, another change from the EDI-2 to EDI-3 was the shift in item scoring from a 4-point format (0, 0, 0, 1, 2, 3) to a 5-point format (0, 0, 1, 2, 3, 4) in the EDI-3 (Garner, 2004). However, previ-ous research (van Strien & Ouwens, 2003) raised its con-cerns about the 4-point format in the EDI-2 and showed that untransformed item scores (1, 2, 3, 4, 5, 6) produced equal or better psychometric properties in terms of reliability, facto-rial integrity, and sensitivity. Hence, this compression of lower scores seems rather arbitrary, and one might question the necessity of the proposed item score transformation.

To summarize, the limited amount of validation studies on the EDI-3 provide some evidence in favor of the 12 new sub-scales, rather weak and mixed evidence for its second-order composite structure, and doubts about the necessity of the item score transformation. To investigate these issues, we study the psychometric properties of the Dutch version of the EDI-3 focusing on the following major aspects: (1) A series of first- and second-order CFA models will be fitted and compared to shed some more light on the factorial validity of the EDI-3. Thereby, we will perform analyses on item- and aggregate-level to ensure comparability with previous research. (2) The comprehensiveness of the composites is examined by using a Schmid–Leiman decomposition to determine which part of the item variance can be ascribed to the composite and which part to subscale specificity. (3) The reliability of subscales and composites will be investigated.

(4) The criterion validity of the EDI-3 will be explored using receiver operating characteristic (ROC) curves testing sensitivity and specificity to distinguish between clinical subjects and a reference group. (5) All analyses will be conducted using both transformed and untransformed item scores. Implications for scoring and the use of the EDI-3 will be discussed. Note that the EDI-3 uses the items of the EDI-2, which were officially published in Dutch (van Strien, 2002) and are used since then.

Methods

Data collection and participants

Data was gathered from patients of two specialized clinics for eating disorders in the Netherlands (Altrecht Eating Disorders Rintveld and GGZ Centraal, Meerkanten). Patients from these two clinics (*n* = 150) were 25.3 years old (*SD* = 7.2) ranging from 18 to 50 years and were predominantly female (*n* = 148, 99%). The sample included 45 patients with anorexia nervosa (AN, 30%), 19 patients with bulimia nervosa (BN, 13%), and 86 patients with an eating disorder not otherwise specified (EDNOS, 57%) who were diagnosed during in-depth unstructured clinical interviews performed by trained psychologists specialized for eating disorders in each clinic. This sample was joined with a similar sample that previously was used in the Dutch validation of the EDI-2 (van Strien & Ouwens, 2003). The latter sample consists of 364 patients (43 were added after publication and 2 were excluded due to forming multivariate outliers, tested with the Mahalanobis distance) aged between 15 and 53 years (*M* = 25.7; *SD* = 6.6 years), and was also predominantly female (*n* = 354, 97.3%). Patient information was given about the kind and amount of eating and purging behavior, weight, and other relevant characteristics to group patients into the main eating disorder categories using the EDI screening list (EDI-SC; see also van Strien & Ouwens, 2003). The first and second author rated all patients with respect to this screening list and had an agreement of 93.4%. The sample included 57 AN patients (16%), 116 BN patients (32%), and 191 EDNOS patients (53%).

Both samples were combined to a joined[1] Dutch clinical sample (*n* = 514) consisting of 102 patients with AN (20%) with a mean body mass index (BMI) of 17.1 (*SD* = 1.8), 135 BN patients (26%) with a mean BMI of 22.1 (*SD* = 5.2), and 277 EDNOS patients (54%) reporting a BMI of 19.3 (*SD* = 5.5).

In addition, and functioning as a nonclinical reference group, Dutch undergraduate female psychology students (*n* = 270) at Tilburg University completed an online version of the EDI-3 in return for course credits. Students were aged between 18 and 40 years (*M* = 20.0; *SD* = 2.7) and reported a mean BMI of 21.6 (*SD* = 2.8).

Statistical Analyses

SPSS Amos (version 18) was used to conduct CFAs with the clinical data at item and aggregated level, testing four models that are schematically presented in Figure 1. [M1]: 12 freely correlating subscales; [M2]: 12 subscales and 5 correlated second-order composites: EDRC, IC, IPC, APC, and OC (Garner, 2004); [M3]: 12 subscales and 2 second-order factors: the *risk composite* (EDRC) and GPMC; and finally [M4]: 12 subscales and 1 general second-order factor comprising all subscales to test the utility of an EDI total score. A Schmid–Leiman variance decomposition was used to determine which part of the item variance can be ascribed to the higher order factor and which part to subscale-specificity (Schmid & Leiman, 1957).

All models were specified starting from the covariance matrix and fitted by means of maximum likelihood. Model fit was evaluated based on commonly recommended goodness-of-fit indices (see, for example, Hu & Bentler, 1999), including the χ^2 of the model fit, the Root Mean Square Error of Approximation (RMSEA), the Tucker-Lewis Index (TLI), the Standardized Root Mean Square Residual (SRMR), and the Bayesian Information Criterion (BIC). Note that because statistical inference in CFA relies on the appropriateness of the sample covariance matrix as a summary of the data, the data were first screened for extreme deviations of normality; no problems were identified.

For assessing reliability, Cronbach's alpha was computed for each subscale and composite reliability (Raykov, 1997) for each second-order composite of the EDI-3. Criterion validity was assessed using Area-Under-the-Curve (AUC) statistics resulting from ROC curves that examine sensitivity and specificity of the EDI-3 subscales in identifying clinical versus nonclinical subjects. AUC scores of .50 indicate no distinctive ability of the subscale to distinguish between both groups, while scores of .80 indicate that with this subscale, 80% of the subjects could be correctly classified (Fawcett, 2006; Mason & Graham, 2002). All analyses were performed using transformed and untransformed item scores. If not stated otherwise, scores are reported for untransformed scores only, in case both scoring patterns led to similar results.

Results

Factorial Validity

Results of the CFAs for both scoring patterns are shown in Table 1. Model comparisons showed that the best fitting model at item level was found in the unrestricted CFA with the 12 correlated subscales (M1) for both scoring patterns. Slightly poorer model fits were found for Garner's second-order factor model with 5 composites (M2). This model also

Table 1. Model Fit and Comparison of the Item-Level Confirmatory Factor Models in the Dutch Clinical Sample.

Model	χ^2	*df*	RMSEA	TLI	SRMR	BIC
Item level						
Untransformed scores: (1,2,3,4,5,6)						
M1	9741	3849	.055	.736	.077	142798
M2	10100	3892	.056	.725	.083	142889
M3	10238	3903	.056	.720	.085	142958
M4	10483	3904	.057	.709	.090	143197
Transformed scores: (0,0,1,2,3,4)						
M1	9368	3849	.053	.743	.075	134120
M2	9729	3892	.054	.731	.081	134212
M3	9879	3903	.055	.726	.083	133684
M4	10109	3904	.056	.715	.089	133911

Note. RMSEA = Root Mean Square Error of Approximation; TLI = Tucker Lewis Index; SRMR = Standardized Root Mean Square Residual; BIC = Bayesian Information Criterion; M1 = 12 correlated subscales; M2 = five second-order composite factors; M3 = two second-order composite factors (Eating Disorder Risk Composite & General Psychological Maladjustment Composite); M4 = one second-order composite factor.

Table 2. Model Fit and Comparison of the Aggregate Subscale Level Confirmatory Factor Models in the Dutch Clinical Sample.

Model	χ^2	*df*	RMSEA	TLI	SRMR	BIC
Subscale level						
Untransformed scores: (1,2,3,4,5,6)						
M2	197	40	.087	.908	.049	13885
M3	466	53	.123	.818	.072	14073
M4	702	54	.153	.719	.087	14303
Transformed scores: (0,0,1,2,3,4)						
M2	196	40	.087	.908	.050	12655
M3	466	53	.123	.815	.073	12844
M4	691	54	.152	.720	.086	13063

Note. RMSEA = Root Mean Square Error of Approximation; TLI = Tucker Lewis Index; SRMR = Standardized Root Mean Square Residual; BIC = Bayesian Information Criterion; M2 = five factors; M3 = two factors (Eating Disorder Risk Composite & GMPC); M4 = one factor.

Table 3. Schmid-Leiman Variance Decomposition for the Second-Order Factor Model 2 Using the Dutch Clinical Sample.

		Average variance explained in an item by			
Composite	Subscale	Composite	Specific	Error	Total
EDRC	DT	.48	.07	.45	1
EDRC	B	.10	.56	.34	1
EDRC	BD	.31	.16	.53	1
IC	LSE	.46	.02	.52	1
IC	PA	.36	.02	.63	1
IPC	II	.24	.09	.67	1
IPC	IA	.22	.09	.69	1
APC	ID	.30	.06	.64	1
APC	ED	.11	.10	.79	1
OC	P	.22	.08	.70	1
OC	A	.20	.08	.72	1
—	MF		.40	.60	1
Overall		.25	.14	.61	1

Note. EDRC = Eating Disorder Risk Composite; DT = Drive for Thinness; B = Bulimia; BD = Body Dissatisfaction; IC = Ineffectiveness Composite; LSE = Low Self-Esteem; PA = Personal Alienation; IPC = Interpersonal Problems Composite; II = Interpersonal Insecurity; IA = Interpersonal Alienation; APC = Affective Problems Composite; ID = Interoceptive Deficits; ED = Emotional Dysregulation; OC = Overcontrol Composite; P = Perfectionism; A = Asceticism; MF = Maturity Fears.

clearly comes forward when directly performing CFAs at the aggregated subscale sum score level (see Table 2). At the aggregated level, this model (M2) is the only one showing acceptable goodness-of-fit with a TLI above .90, a RMSEA around .08, and an SRMR below .08. However, two key problems emerged when examining the factor structure of the EDI-3.

First, the subscales LSE and PA (*r* = .953) showed an extremely high inter-factor correlation, which caused so-called Heywood cases in all second-order models (i.e., error variances equal to 0, combined with extremely large standard errors, and communalities equal to or larger than 1). Second-order factor loadings for both subscales had to be constrained to be equal to avoid this. The empirical underidentification problem that is outlined here is a clear signal of overfactorization as there is almost no information in the

data to differentiate between the two subscale factors LSE and PA.

Second, all models show considerable misfit at the item level as for instance shown in the substandard TLI. The cause of this misfit is partially due to subsets of items that show correlated errors (e.g, Items 79 and 83 [ED], 13 and 44 Perfectionism, and Items 22 and 39 [MF]) and due to item-factor misallocation (e.g., Item 35 [MF] yet possibly PA; Item 34 [II] yet both possibly IA; and Item 26 [ID] yet possibly II) or badly functioning items as shown by very low factor loadings (< .20; for example, Items 47 [BD] and 81 [ED]). Note that preliminary data screening also showed that one item (item 72 [ED]) showed uncorrectable skewness (2.52/3.51) and excess kurtosis (6.17/12.43) for untransformed and transformed scores, respectively. It would be recommended to eliminate this item from further analyses (see Curran, West, & Finch, 1996). Hence, the structure of the EDI-3 at the item level should be further refined. Note that analyses based on the recommended transformed scores and untransformed item scores lead to similar results in the factor analyses.

To shed some more light on the model specification of the EDI-3, we decomposed the variance of each item into three parts: (i) variance explained by the second-order factor (i.e., variance shared by first-order factors), (ii) variance assigned to the specific subscale, and (iii) the remaining unexplained error variance. The results of this so-called Schmid–Leiman decomposition are summarized in Table 3, presenting the

Table 4. Reliability Results for the Eating Disorder Inventory-3 Subscales and Composites in the Dutch Clinical Sample.

Subscale	Transformed scores		Untransformed scores	
	M (SD)	α	M (SD)	α
DT	20.4 (7.1)	.88	34.1 (17.6)	.88
B	13.1 (9.3)	.92	27.5 (11.2)	.93
BD	26.8 (9.8)	.89	46.1 (10.4)	.88
LSE	14.4 (5.3)	.85	26.3 (5.5)	.84
PA	14.6 (5.5)	.79	28.3 (5.7)	.79
II	11.3 (5.1)	.75	24.8 (5.7)	.76
IA	11.5 (5.1)	.74	24.9 (5.7)	.74
ID	17.0 (7.1)	.81	34.5 (7.7)	.81
ED	8.1 (4.5)	.64	21.8 (5.4)	.65
P	10.9 (5.0)	.73	22.3 (5.5)	.74
A	12.7 (5.1)	.71	25.9 (5.6)	.71
MF	11.7 (6.6)	.84	26.6 (7.5)	.84
Composite	M (SD)	Rel.	M (SD)	Rel.
EDRC	60.3 (20.6)	.70	107.7 (23.0)	.68
IC	29.0 (10.3)	.89	54.6 (10.8)	.88
IPC	22.8 (09.1)	.77	49.7 (10.2)	.75
APC	25.0 (10.2)	.75	56.3 (11.4)	.73
OC	23.5 (08.7)	.76	48.2 (09.6)	.74
GPMC	112.0 (34.2)	.87	235.4 (37.6)	.87

Note. DT = Drive for Thinness; B = Bulimia; BD = Body Dissatisfaction; LSE = Low Self-Esteem; PA = Personal Alienation; II = Interpersonal Insecurity; IA = Interpersonal Alienation; ID = Interoceptive Deficits; ED = Emotional Dysregulation; P = Perfectionism; A = Asceticism; MF = Maturity Fears; EDRC = Eating Disorder risk Composite; IC = Ineffectiveness Composite; IPC = Interpersonal Problems Composite; APC = Affective Problems Composite; OC = Overcontrol Composite; GPMC = General Psychological Maladjustment Composite.

Table 5. AUC Results From Receiver Operating Characteristic Curve Analyses Distinguishing Clinical Versus Nonclinical Sample Based on the Eating Disorder Inventory-3 Subscale Scores.

Subscale	AUC
DT	.903
B	.776
BD	.849
LSE	.906
PA	.899
II	.697
IA	.743
ID	.901
ED	.779
P	.768
A	.902
MF	.678

Note. AUC = Area-Under-the-Curve; DT = Drive for Thinness; B = Bulimia; BD = Body Dissatisfaction; Rel.= Reliability; LSE = Low Self-Esteem; PA = Personal Alienation; II = Interpersonal Insecurity; IA = Interpersonal Alienation; ID = Interoceptive Deficits; ED = Emotional Dysregulation; P = Perfectionism; A = Asceticism; MF = Maturity Fears.

Composite reliabilities were also similar across scoring patterns, ranging between .68 (EDRC) and .89 (IC). In comparison to the other composites, the EDRC showed a rather low reliability, although each of the included subscales showed a large reliability. Separate testing in the clinical groups showed particularly low EDRC reliability in the BN group (.59/.53). This adds up to the previous observation that items in the Bulimia subscale only contain 10% of variance explained by the composite, yet 56% of variance due to the specific subscale. Therefore, its separate use is recommended.

Criterion Validity: Sensitivity and Specificity

ROC curve analyses were performed regarding the EDI-3's ability to distinguish between clinical versus nonclinical cases. The AUC statistics (Table 5) indicated good to excellent results for the 12 subscales. Especially, the results for the subscales LSE, DT, A, ID, and PA (.906, .903, .902, .901, and .899, respectively) suggest rates about 90% to correctly identify the subjects as clinical versus nonclinical. These excellent rates were followed by several good rates ranging from .743 (IA) up to .849 (BD). The lowest discrimination was achieved with the subscales II and MF with values of .697 and .678, respectively, which have to be interpreted as poor. The latter subscale also shows rather low correlations with the other subscales in the results of the factor analyses.

Discussion

Testing the psychometric properties of the EDI-3 is of great importance, because (any version of the) EDI is frequently used in clinical practice and research, but independent

average variances per part (i.e., composite, subscale, and error) across the 12 EDI-subscales. Note that the three parts sum up to the total variance and scores are reported for the untransformed scores. Given the large subscale-specific explained variance of the Bulimia subscale, we recommend to remove this subscale from the EDRC and use it separately. Furthermore, we advice to no longer distinguish between subscales LSE and PA, as they merely provide 2% of subscale-specific variance and instead directly talk about the IC as one subscale. For the remaining two-scale composites, the subscales still add about 10% specific variance to the variance explained by their composite that we considered meaningful. Note that subscales with high residual variance, such as ED or A should be candidates for item revision and refinement (see also Table 3).

Reliability

Reliability analyses (see Table 4) yielded acceptable to excellent results for the 12 subscales introduced by Garner. Cronbach's alphas were found similar for both scoring patterns and ranged between .64 (Asceticism) and .93 (Bulimia).

research *about* the most recent version is scarce. Therefore, the aim of this study was fivefold including the examination of the factorial validity, reliability, and sensitivity and specificity of the Dutch version of the EDI-3 in a large sample of eating disordered patients using transformed and untransformed item scores.

First, factorial validity has been found to be the best for the model with 12 freely correlating factors. However, it should be noted that the structure of the EDI-3 at the item level can (and perhaps should) be further refined, because model fit was for instance impaired by the presence of correlated item errors and misallocated items. At the aggregate level, support was found for a second-order model with two second-order factors representing an eating-disorder specific and psychological composite that supports Garner's initial validation and the recent Spanish study (Elosua & López-Jáuregui, 2011). However, such aggregate models do not examine the fit of each item to the subscales, and hence whether they are good indicators for the construct intended to be measured. Because we found serious problems at the item level, we pled for further validation studies or revisions of the EDI-3 that target the item-level as it has been done by Clausen and colleagues (2011) and Garcia-Grau and colleagues (2010). Their findings also suggest problems regarding the item allocation to the suggested subscales, and we need further studies to gain more clarity in what is exactly measured by the different factors.

Second and further refining aspects for the factorial validity, we decomposed the items' variances according to their subscale specificity, composite, and error variance. The results indicated that the EDI-3 factor structure might be subject to overfactorization and contains more theoretical common factors than there are supported by the data. For instance, the two subscales LSE and ID are empirically indistinguishable and are better taken together under the denominator of their second-order composite. Furthermore, it was shown that the Bulimia subscale did not contribute much to its EDRC composite and was largely a specific source of information. We would therefore recommend the use of this subscale on its own in practice, instead of as part of a composite. It was also confirmed that especially the ED subscale suffered the most from model misfit at the item level.

Third, reliability analyses of the 12 subscales indicated good to excellent results with only one score below .7 for the ED subscale. Hence, the overall integrity of this scale has to be questioned given both its unconvincing factorial and reliability results. Composite reliabilities showed good results. However, the risk composite yielded the lowest scores with .67/.68 (transformed/untransformed scores). This again supports the separate usage of the Bulimia subscale, which is also reasonable considering theoretical aspects of the eating disorder categories: Bulimic symptoms are applicable to some, but not all, eating disorder categories, that is, to bulimia and cases of purging anorexia but not restrictive anorexia and not completely to binge eating disorder. Hence, bulimic

symptoms are not overall applicable to be joined to a general EDRC, while the other two subscales (BD and DT) are applicable to almost all clinical presentations of eating disorders.

Fourth, sensitivity and specificity showed good to excellent results using Garner's 12 subscales, but a poor score for the subscale MF suggests a poor discriminative ability between clinical versus nonclinical subjects. This scale tries to capture etiological issues of eating disorders instead of current symptoms. In addition, this etiological aspect is predominantly relevant to anorexia (Wonderlich, 2002) and therefore its discriminative ability for the other eating disorder diagnoses may be weak, although more research in this respect is needed.

Fifth, the usage of transformed item scores seems redundant, because both scoring patterns showed similar distributions and all investigated psychometric aspects showed similar results and comparable fit indices. Hence, our results are in line with the conclusions of van Strien and Ouwens (2003), who identified the use of untransformed item scores to work equally well in clinical samples and not harming validity.

Eventually, we want to address certain limitations here. The participants in the clinical sample were not diagnosed using a clinical structured interview and diagnostic techniques differed in the two initial subsamples. However, in the first clinical subsample, conform common clinical practice, diagnoses were given during in-depth interviews and by experienced clinicians. In the second subsample, diagnoses were allocated using the EDI-SC, and a high agreement rate of almost 94% was achieved assuring accurate diagnoses. Furthermore, one might question the suitability of the nonclinical convenience sample used for the ROC analyses, because female students in their early twenties may also be more concerned with their eating and show symptoms of disordered eating. However, bearing this in mind, the discriminative ability of the EDI-3 is even more impressive as results were excellent successfully distinguishing between the clinical versus student sample.

In conclusion, the 12-factor model of the EDI-3 received limited support according to factorial validity and excellent results concerning reliability as well as sensitivity and specificity. The EDI-3 is a very comprehensive measure trying to tap numerous aspects of eating disorders, but the factorial integrity leaves much room for improvement and we call for more psychometric research about this instrument. Till then, it should be used with caution in the clinical practice. In particular, the usage of the EDRC is not recommended as the Bulimia subscale should be handled as a separate source of information. In addition, overfactorization seemed to be a general problem of the EDI-3. Therefore, subscales tapping on interpersonal problems and alienation may not be seen as diverse sources of information as they correlate very highly. The added value of the subscale MF should also be regarded critically in the clinical practice, as it proved little valuable for the discriminative validity and showed only low correlations with all other EDI-3 subscales. Clinicians may

critically consider the value and sense of this subscale as it does not seem relevant for the development and pathology for all types of eating disorders. However, all other EDI-3 subscales showed a good discriminative validity, which gives us confidence that the EDI-3 may prove to be a valuable tool for screening.

Declaration of Conflicting Interests

Funding

Note

1. On top of the similarities in context and setting, preliminary statistical analyses showed that both samples were metric invariant across the Eating Disorder Inventory–3 subscales (with some caution for the items 10, 27, 42, and 87).

References

Clausen, L., Rosenvinge, J. H., Friborg, O., & Rokkedal, K. (2011). Validating the Eating Disorder Inventory-3 (EDI-3): A comparison between 561 female eating disorder patients and 878 females from the general population. *Journal of Psychopathology and Behavioral Assessment*, *33*, 101-110. doi:10.1007/s10862-010-9207-4

Curran, P., West, S., & Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29. doi:10.1037/1082-989X.1.1.16

Elosua, P., & López-Jáuregui, A. (2011). Internal Structure of the Spanish Adaptation of the Eating Disorder Inventory-3. *European Journal of Psychological Assessment*, *28*, 25-31. doi:10.1027/1015-5759/a000087

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861-874. doi:10.1016/j.patrec.2005.10.010

Garcia-Grau, E., Fuste, A., Mas, N., Gomez, J., Bados, A., & Saldana, C. (2010). Dimensionality of three versions of the Eating Disorder Inventory in adolescent girls. *European Eating Disorder Review*, *18*, 318-327. doi:10.1002/erv.995

Garner, D. M. (2004). *EDI-3 Eating Disorders Inventory-3: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, *128*, 2145-2166. doi:10.1256/003590002320603584

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*, 173-184. doi:10.1177/01466216970212006

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53-61. doi:10.1007/BF02289209

van Strien, T. (2002). *Eating Disorder Inventory II— Nederlandse Versie (EDI-II-NL)* [Dutch Version]. Amsterdam, The Netherlands: Swets Test Publishers.

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. doi:10.1027//1015-5759.19.1.66

Wonderlich, S. A. (2002). Personality and eating disorders. In C. G. Fairburn & K. D. Brownell (Eds.), *Eating disorders and obesity* (pp. 204-209). New York, NY: Guilford Press.

Author Biographies

Vicky Lehmann is a clinical psychologist and currently obtaining her PhD degree at the University Medical Center Groningen, The Netherlands. Her research interest focused on psycho-pathological aspects in eating and psychosomatic disorders. Currently, she extended her focus to psychological long-term effects of childhood cancer.

Machteld A. Ouwens is a clinical psychologist and cognitive behavioral therapist at GGzBreburg where she is treating patients with mood and anxiety disorders. At the time of the present study, she was working as an assistant professor at Tilburg University, department of Clinical Psychology, conducting research on attachment, emotion regulation, and disturbed eating behavior.

Johan Braeken is an assistant professor in the Research Methodology group at Wageningen University and Research centre. As a psychometrician he is interested in all aspects of assessment, evaluation, and testing, with a specific passion for the development and application of latent variable models in a measurement context.

Unna N. Danner is a senior researcher at Altrecht Eating Disorders Rintveld. Her areas of research extend to neuropsychological functioning of patients with eating disorders. She has a particular interest in the influence of affective problems on decision making processes and the (cost)effectiveness of cognitive remediation therapy for patients with severe anorexia nervosa.

Annemarie A. van Elburg is a professor of eating disorders at the University of Utrecht, member of the Utrecht Research Group for Eating disorders and Medical Director of Rintveld, center for eating disorders of Altrecht Mental Health Inst. Her research expertise includes translational research in eating disorders, treatment studies and cost effectiveness studies.

Marrie H. J. Bekker holds the main chair of Clinical Psychology at Tilburg University. Her core research domains are autonomy, attachment, emotion-regulation, psychopathology, as well as mental health (care) and ethnic and gender-related diversity.

Annette M. Breurkens is a psychotherapist with an extensive experience in treating eating disordered patients in a clinical setting at GGz Centraal in Harderwijk and in her own private practice. She also teaches postgraduates as part of their curriculum to become clinical psychologists or psychotherapists.

Tatjana van Strien is endowed professor Psychology of eating styles at the Free University in Amsterdam and is also associated with the University of Nijmegen. Her teaching and research interests focus on the psychology of dieting and overeating. She has authored the Dutch Eating Behaviour Questionnaire (DEBQ).