RESEARCH ARTICLE

Internal Structure and Measurement Invariance of the Dutch Eating Behavior Questionnaire (DEBQ) in a (Nearly) Representative Dutch Community Sample

Juan Ramón Barrada[1]* , Tatjana van Strien[2,3] & Ausiàs Cebolla[4,5]

[1]Facultad de Ciencias Sociales y Humanas, Universidad de Zaragoza, Spain
[2]Institute for Gender Studies and Behavioural Science Institute, Radboud University, The Netherlands
[3]Faculty of Earth and Life Sciences, Institute of Health Sciences, VU University Amsterdam, The Netherlands
[4]Universitat de València, Spain
[5]CIBER Fisiopatología Obesidad y Nutrición (CB06/03), Instituto Carlos III, Spain

Abstract

The Dutch Eating Behavior Questionnaire is a widely used instrument for assessment of emotional, external and restrained eating. The aim of the present study is to (i) analyse its internal structure using exploratory structural equation modelling; (ii) to assess its measurement invariance with respect to sex, BMI, age and level of education; and (iii) to evaluate the relations of the factors with these variables. Except that women were slightly over-represented, the sample ($n = 2173$) closely followed the sociodemographic characteristics of the overall Dutch population. The three theoretical factors that emerged from the analysis were in close correspondence with the three scales for emotional, external and restrained eating. Only two items (item 3 — 'desire to eat when nothing to do…' and item 21 — 'resist delicious food...') presented problematic loadings. The questionnaire showed satisfactory measurement invariance, and expected patterns of mean differences and relations were found. All in all, the results highlight the adequate psychometric properties of the Dutch Eating Behavior Questionnaire. Copyright © 2016 John Wiley & Sons, Ltd and Eating Disorders Association.

Introduction

Eating styles refer to a complex interplay amongst physiological, psychological, social and genetic factors that influence food preferences and quantity of food intake (Grimm & Steinle, 2011), and show strong associations with obesity and eating disorders (Cebolla, Barrada, van Strien, Oliver, & Baños, 2014). Although eating styles can be assessed with various scales, the Dutch Eating Behavior Questionnaire (DEBQ; van Strien, Frijters, Bergers, & Defares, 1986) is the only questionnaire that simultaneously covers three main eating styles and that has been developed in community samples.

The DEBQ eating styles are based on three main psychological theories on overeating: psychosomatic theory (Bruch, 1973), externality theory (Schachter, 1971) and restraint theory (Polivy & Herman, 1985). Psychosomatic theory focuses on emotional eating (eating in response to negative emotions such as depression and discouragement) as an atypical response to distress. Externality theory emphasizes eating in response to food-related stimuli, such as the sight, smell and taste of food, regardless of the internal state of hunger and satiety. Restraint theory focuses on the possible psychological side effects of dieting, specifically the disinhibition effect: overeating by dieters when their cognitive resolve to eat less than desired is abandoned.

The DEBQ: development and psychometric properties

The 33 items on the DEBQ represent three separate scales: (i) *Emotional Eating*, with items such as 'Do you have the desire to eat when you are irritated?', contains 13 items; (ii) *External Eating*, with items such as 'Do you eat more than usual when you see others eating?', includes 10 items; and (iii) *Restraint*, with items such as 'Do you deliberately eat less in order not to become heavier?', has 10 items. In the current Dutch version, the original version, all items are direct, except for item 21, which is reverse-scored ('Do you find it hard to resist eating delicious foods?'). In this version, the response format is a Likert-type scale ranging from 1 = *seldom* to 5 = *very often*.

The English version of the DEBQ (van Strien, 2002) and those adaptations of this version incorporate two modifications to the Dutch original: (i) the item response *seldom* was renamed *rarely*, whilst keeping the five-point response options; and (ii) item 21 was reformulated so it is now direct.

The psychometric properties of the DEBQ have been evaluated with its original version (van Strien et al., 1986; $n = 616$ for the final study) and several translated versions: English (Wardle, 1987; $n = 188$), French (Lluch et al., 1996; $n = 82$), Italian (Dakanalis et al., 2013; $n = 990$), Spanish (Cebolla et al., 2014; $n = 647$) and Turkish (Bozan, Bas, & Asci, 2011; $n = 508$). All translated versions included the described modifications of the original DEBQ.

Some common results can be found across all these studies. The structure of the DEBQ seems to be well explained by the three theoretical components, although with the option of subdividing the emotional dimension into two separate factors (van Strien et al., 1986): a factor, labelled *Diffuse Emotions Eating* (e.g. 'Do you have a desire to eat when you have nothing to do?') and a factor tapping clearly labelled emotions (e.g. 'Do you have a desire to eat when you are irritated?'). The *Emotional Eating* and *Restraint* factors have somewhat similar high average factor loadings, with the loadings for *External Eating* being a bit lower. The same can be said about the internal consistency. The factors with more shared variance are *Emotional* and *External Eating*, whilst the least related factors are *Restrained* and *External Eating*.

All the previous studies that have analysed the internal structure of the DEBQ have used a nonrepresentative convenience sample. Furthermore, almost all of them have used nonoptimal psychometric approaches, with the Italian (Dakanalis et al., 2013) and Spanish (Cebolla et al., 2014) versions being noteworthy exceptions. We will point out the main limitations of those studies (excluding Dakanalis et al. and Cebolla et al.). Only exploratory factor analysis (EFA) with Varimax (orthogonal) rotation as the analytical technique was used. None of the studies, except the Turkish version (Bozan et al., 2011), reported the extraction method. Furthermore, the sample sizes for the English (Wardle, 1987; $n = 188$) and French versions (Lluch et al., 1996; $n = 82$) can be considered somewhat small for analytical purposes.

In the Italian version (Dakanalis et al., 2013), two separate EFA ($n = 495$) and confirmatory factor analysis (CFA; $n = 495$) were carried out over the 33 items of the DEBQ. For both analyses, the results supported the theoretical distribution of items in three correlated factors. Importantly, the results indicated that the DEBQ scores offered measurement invariance (equal scores can lead to equal interpretation across groups) with respect to sex, BMI status and age. With the Spanish version (Cebolla et al., 2014), inter-item correlations were modelled with an exploratory structural equation model (ESEM; Asparouhov & Muthén, 2009) and a CFA. In the ESEM technique, the uniqueness of the items belonging to the *Diffuse Emotions* category was allowed to correlate; the same for three item pairs with high content similarity. Although the model fit was satisfactory, item 3 ('Do you have a desire to eat when you have nothing to do?') and item 28 ('Do you have a desire to eat when you are bored or restless?') presented problematic cross-loadings in the *Emotional* and *External Eating* dimensions. When considering the 33 items, the ESEM approach provided a better fit to the data.

Purposes of the study and expected results

To date, just two studies have examined the internal structure of the DEBQ using a modern psychometric approach (Cebolla et al., 2014; Dakanalis et al., 2013), and their results are not completely convergent. As already stated, none of the previous studies have used a representative sample. Therefore, there are still some open questions about the factorial structure of the DEBQ. The purposes of the present study were:

1. To analyse the internal structure of the DEBQ.
2. Following the line initiated by Dakanalis et al. (2013), to evaluate the measurement invariance of the DEBQ considering different relevant subgroups. In addition to sex, BMI and age, we will also assess the measurement invariance in regard to level of education.
3. If evidence supporting measurement invariance (or partial invariance) is found, we will analyse the relationship between these variables and the eating styles.

Method

Participants

Data were collected in 2006 and 2009 by a panel service agency in two different cohorts of Dutch people over 15 years with the aim that the samples were representative for The Netherlands in regard to sex, level of education, ethnic origin and region ($n = 2173$; van Strien, Herman, & Verheijden, 2009, 2012). Data of 63.0% of the participants in the present study had been collected in 2006. The main descriptives of the present sample are (i) with respect to sex, women comprised 57.6% of the sample; (ii) BMI (calculated from self-reported height and weight) ranged from 13.68 to 65.75 ($M = 25.54$, $SD = 4.82$), with 2.7% underweight, 48.5% normal weight, 33.8% overweight and 15% obese; (iii) the mean age was 39.6 [$SD = 13.9$, range (16–77)]; and (iv) with respect to level of education, 25.3% of the sample had a lower level (primary school, lower vocational education or lower general secondary education), 44.1% had a middle level of education (intermediate vocational education, higher general secondary education or pre-university education) and 30.6% had a higher level of education (higher vocational education or academic training). All the participants had complete data on the DEBQ scales. Other variables had a small number of missing data, never greater than seven participants, which was considered trivial. As can be seen in Table 1, the present sample follows to an important degree the sociodemographic characteristics of the overall Dutch population. The main departure from representativeness is with respect to the percentage of women and men, with an over-representation of the women. The close resemblance of the sample to the overall Dutch population led the Dutch Committee on Tests and Testing (COTAN) to rate the norms of the DEBQ (which requires representativeness of the sample for the Dutch population Evers, Sijtsma, Lucassen, & Meijer, 2010) with the highest evaluation, namely 'good' (COTAN, 2013). Following the COTAN, but also considering the discrepancy between the proportion of women–men in our sample and the Dutch population, we consider our sample as nearly representative. Further information on the sample characteristics can be found in van Strien et al. (2009, 2012).

Data analysis

First, we studied the internal structure of the scale with an ESEM approach (Asparouhov & Muthén, 2009). ESEM is a technique that, unlike CFA, permits that all items load into all factors, and unlike EFA, permits the correlation between item uniqueness, and to test measurement invariance. Given the results of Cebolla et al. (2014), who found that the ESEM approach provided a

Table 1 Sample and Dutch population distribution in main sociodemographics

		Sample	Dutch population[†]
Region[‡]	Three big cities	14.3 %	16.6%
	West	28.4%	29.2%
	North	10.7%	10.0%
	East	21.8%	20.9%
	South	24.8%	23.3%
Ethnicity	Autochthonous	79.4%	79%
	Allochthonous	20.6%	21%
Education	Low	25.3%	30%
	Medium	44.1%	42%
	High	30.6%	28%
sex	Male	42.4%	49.5%
	Female	57.6%	50.4%
Weight status	Overweight + obese	48.8%	48.3%
	Obese	15%	12%
age	Mean age	39.63 years	39 years (aprox.)

[†]According to the Central Bureau for Statistics (http://www.cbs.nl).

[‡]The regions are the three big cities (Amsterdam, The Hague and Utrecht), west (Utrecht, North and South Holland), north (Groningen, Friesland and Drente), east (Overijssel, Gelderland and Flevoland) and south (Zeeland, Noord Braband and Limburg) parts of The Netherlands.

better fit to the data and the greater flexibility of the ESEM, we considered ESEM the preferred technique for the present study. Models were analysed using robust weighted least squares.

Goodness-of-fit in all derived models was assessed with the common cut-off values for the fit indices (Hu & Bentler, 1999): We consider whether the comparative fit index (CFI) and Tucker–Lewis index (TLI) have values greater than 0.95, and whether the root mean square error of approximation (RMSEA) is less than 0.06. Localized areas of strain will be assessed with modification indices. We tested several models until we derived a final model with satisfactory fit. As we have noted, van Strien et al. (1986) showed the possibility of splitting the *Emotional Eating* dimension into two factors. We consider that correlated uniqueness amongst the items tapping diffuse emotions was a better option than a possible new factor because (i) that new factor would have no theoretical advantage with respect to the general *Emotional Eating* factor; and (ii) the correlation between both factors found in van Strien et al. (1986) indicated a poor discriminant validity between them. So, in case that the initial three-factor model does not show satisfactory fit, we will test a model where those correlated uniquenesses are included. After defining the internal structure of the scale, we evaluated the reliability of the scores for each dimension. The reliabilities of the sum of the observed scores were computed with coefficient omega (McDonald, 1999).

We also carried out a factor invariance study, splitting the sample by the variables in the Italian study by Dakanalis et al. (2013; sex, BMI and age) and, additionally, level of education. For testing the invariance, the equality (or minimal difference) of the fit between consecutive models is evaluated. First of all, it was verified that the model fitted each group separately. Second, we tested the equality of form. In the context of ESEM, this involves fixing the number of factors and pairs of correlated uniqueness. Third, we tested the equality of thresholds and factor loadings across groups

(Muthén & Muthén, 1998–2012). We considered these restrictions to be satisfactorily met if the decrease in CFI was lower than 0.01, and RMSEA and TLI remained constant or increased. For continuous variables, such as BMI and age, an analysis was performed, dividing the sample in three roughly equal groups by computing the tertiles. The cut-off points for BMI were 23.15 and 26.73, and for age, they were 32 and 41.

Once measurement invariance had been justified, we studied the relation between these key variables and eating styles. To do so, we used a multiple indicator multiple cause (MIMIC) approach. For level of education, two dummy variables were created to compare the middle category with the others. With the MIMIC model, we could evaluate the mean differences (for categorical variables) or the expected effect of the increment of one standard deviation (for continuous variables) on the standardized scores of the factors, whilst controlling for the rest of the variables. All the analyses were performed with M*plus* 7 (Muthén & Muthén, 1998–2012).

Results

Internal structure of the DEBQ

The fit of the different models can be seen in Table 2. According to conventional cut-off values, this model (M1, where M denotes 'model' in this and next cases) failed to fit. Following the results from van Strien et al. (1986) and Cebolla et al. (2014), in the next tested model (M2) the item uniquenesses of all the items without clearly labelled emotions were allowed to correlate. All the correlated uniquenesses were statistically significant, with a mean of 0.287. Although the fit of M2 was improved with respect to M1, it could still not be considered a well fitted model.

We inspected the modification indices of M2 in order to detect localized areas of strain. Modification indices as high as 787.12 were found, in this case, indicating the appropriateness of freeing the correlation between the uniquenesses of items 2 and 6. These two items can be considered the same concept with two slightly different wordings. An iterative process was begun, where the greatest modification index indicated the need for correlated uniquenesses to be introduced in the next model. In this way, we allowed items 2 and 6 to correlate ($r = .58$), as well as items 4 and 19 ($r = .44$) and items 26 and 29 ($r = .37$; all $ps < .001$ and r values taken from M3). We finished the process when the next two items to be included (items 25 and 30) were not as clearly overlapping as the previous pairs. M3 is the model with correlated uniquenesses between the diffuse emotion items and those items located through modification indices. We retained this as the final model (although its RMSEA is slightly higher than conventional cut-off values; RMSEA = 0.066 vs. RMSEA ≤ 0.060).

Item loadings for M3 are shown in Table 3. Several points about the results deserve attention. First, two problematic items were found. Item 21, the reverse-keyed item, was found to be a poor indicator of all dimensions of eating styles, as its highest loading was as low as 0.13. Item 3, although its highest loading was in its intended factor, presented a loading of almost the same size in another factor. Second, for the target loadings (loadings in the factor that the items are intended to measure), and not considering items 3 and 21, as expected, those from *External Eating* ($M = 0.64$) were lower than those from *Emotional Eating* ($M = 0.82$) or *Restraint* ($M = 0.77$). These values can be considered

Table 2 Goodness of fit indices for the different models

Models		$\chi^{2\dagger}$	df	RMSEA	TLI	CFI	ΔRMSEA	ΔTLI	ΔCFI
M1	Three factors	7295.41	432	0.086 [0.084, 0.087]	0.943	0.953			
M2	Three factors — CU DIFFUSE	6267.94	426	0.079 [0.078, 0.081]	0.960	0.951			
M3	Three factors — CU DIFFUSE and RESP	4371.48	423	0.066 [0.064, 0.067]	0.966	0.973			
	Measurement invariance								
	By sex — CU DIFFUSE and RESP								
M4	Woman	2831.46	423	0.068 [0.065, 0.070]	0.961	0.969			
M5	Men	1847.52	423	0.060 [0.058, 0.063]	0.969	0.975			
M6	Equal form	4715.40	846	0.065 [0.063, 0.067]	0.964	0.971			
M7	… and equal loadings and thresholds	3531.37	1032	0.047 [0.046, 0.049]	0.981	0.981	−0.018	0.017	0.010
	By BMI — CU DIFFUSE and RESP								
M8	Lower (BMI ≤ 23.15)	1679.74	423	0.064 [0.061, 0.067]	0.969	0.975			
M9	Middle (23.15 < BMI ≤ 26.73)	1325.25	423	0.054 [0.051, 0.058]	0.973	0.978			
M10	Higher (BMI > 26.73)	1892.99	423	0.069 [0.066, 0.072]	0.962	0.970			
M11	Equal form	4730.71	1269	0.061 [0.060, 0.063]	0.969	0.975			
M12	… and equal loadings and thresholds	3715.03	1641	0.042 [0.040, 0.044]	0.986	0.985	−0.019	0.017	0.010
	By age — CU DIFFUSE and RESP								
M13	Younger (age ≤ 32)	1942.35	423	0.068 [0.065, 0.071]	0.961	0.969			
M14	Middle (32 < age ≤ 41)	1442.13	423	0.060 [0.056, 0.063]	0.969	0.976			
M15	Older (age > 41)	1467.26	423	0.059 [0.056, 0.062]	0.975	0.980			
M16	Equal form	4727.10	1269	0.061 [0.059, 0.063]	0.969	0.975			
M17	… and equal loadings and thresholds	4532.88	1641	0.049 [0.048, 0.051]	0.980	0.980	−0.012	0.011	0.005
	By education — CU DIFFUSE and RESP								
M18	Lower	1180.68	423	0.057 [0.053, 0.061]	0.973	0.978			
M19	Middle	2154.35	423	0.065 [0.063, 0.068]	0.968	0.975			
M20	Higher	1517.28	423	0.062 [0.059, 0.066]	0.969	0.975			
M21	Equal form	4922.82	1269	0.063 [0.061, 0.065]	0.969	0.975			
M22	… and equal loadings and thresholds	3562.74	1641	0.040 [0.038, 0.042]	0.987	0.987	−0.023	0.018	0.012
M23	MIMIC model	4828.91	573	0.059 [0.057, 0.060]	0.959	0.966			

Notes: Three factors, *Emotional Eating*, *External Eating* and *Restraint*; four factors, *Emotional Eating* (items with a clearly labelled emotion), *External Eating*, *Restraint* and *Diffuse emotional Eating*; CU, correlated uniqueness; DIFFUSE, diffuse emotions; RESP, model respecifications after inspection of modification indices; df, degrees of freedom; RMSEA, root mean square error of approximation with its 90% confidence interval; TLI, Tucker–Lewis index; CFI, comparative fit index; Δ, increment in fit index with respect to previous model.

[†]All p-values for the Chi-square test were < .001.

high. Third, the nontarget loadings (loadings in factors that the item was not intended to measure) in absolute values (again, not considering items 3 and 21) had a low mean value ($M = 0.06$) ranging from 0.00 to 0.29. When an ESEM model without items 3 and 21 was tested, the fit (RMSEA = 0.068, TLI = 0.967, CFI = 0.974) was basically equal to that of the ESEM model with all the items (M3).

The inter-factor correlations followed the expected pattern, given the results of previous studies. The most closely related dimensions were *Emotional* and *External Eating*, whilst the dimensions that shared lower variance were *Restraint* and *External Eating*.

Measurement invariance and relation with external variables

We first checked the model fit for each group separately. As can be seen in Table 2, the fit was satisfactory for all the groups produced by splitting the sample according to the relevant variables, that is, sex (M4 and M5), BMI (M8, M9 and M10), age (M13, M14 and M15) and level of education (M18, M19 and M20). Then, we tested equality of form. The fit was satisfactory in this step for all the segmentations of the sample (M6, M11, M16 and M21).

Lastly, we tested the equality of thresholds and factor loadings (M7, M12, M17 and M22). For all the cases, model fit did not even decrease when constraints were imposed, instead it improved, probably a result of model parsimony, as a lower number of parameters had to be estimated. Considering this, we could establish the measurement invariance of the DEBQ with respect to the tested variables.

The results of this MIMIC analysis are shown in Table 4. The pattern of results basically followed what could be expected given the theory and previous results. First, men had lower means than women on *Emotional Eating* and *Restraint*, whilst both groups had basically equivalent means on *External Eating*. Second, the three scales' scores increased with BMI, with *External Eating* being the dimension least related to BMI. Third, *Emotional* and *External Eating* decreased with age, whilst restraint behaviours tended to increase. Fourth, those participants with a lower level of education presented lower mean scores in all three factors, whilst differences between those with middle and higher education were negligible or nonsignificant. Fifth, the percentage of explained variance for all factors was medium–low for all dimensions, with *External Eating* being the least well explained.

Table 3 Factor loadings and inter-factor correlations for ESEM (M3) model

Item wording[†]	Emot	Restr	Exter
01. Desire to eat when irritated…	**0.84**	0.03	−0.02
02. Eat more than usual when food is tasty…	0.01	−0.05	**0.66**
03. *Desire to eat when nothing to do…*	**0.42**	0.02	0.37
04. Eat less than usual after weight gain	−0.03	**0.75**	0.02
05. Desire to eat when feeling depressed…	**0.86**	0.08	0.00
06. Eat more than normal when food is good…	−0.06	−0.02	**0.73**
07. Reject food or drinks because of worry about weight…	0.09	**0.77**	−0.13
08. Eating when you feel lonely…	**0.65**	0.07	0.16
09. Desire to keep eating when delicious…	−0.12	0.03	**0.82**
10. Desire to eat when somebody lets you down…	**0.87**	0.02	−0.01
11. Eat less during meal times…	0.09	**0.67**	0.10
12. Eat delicious food immediately	0.08	−0.13	**0.50**
13. Desire to eat when angry…	**0.86**	−0.03	0.05
14. Watch what you eat…	0.01	**0.78**	−0.17
15. Desire to eat something delicious…	0.03	−0.01	**0.61**
16. Desire to eat when feeling unpleasant…	**0.82**	−0.03	0.09
17. Eat foods that are slimming…	0.01	**0.77**	−0.04
18. Desire to eat when others are eating…	0.00	0.01	**0.75**
19. Eating less after eating too much…	−0.02	**0.76**	0.01
20. Desire to eat when anxious…	**0.91**	0.01	−0.03
21.*Resist delicious food…*	−0.01	0.04	**0.13**
22. Eat less deliberately…	−0.05	**0.90**	0.01
23. Desire to eat when things go against you…	**0.95**	0.01	−0.03
24. Desire to buy food when passing snack bar…	0.01	−0.17	**0.70**
25. Desire to eat when upset…	**0.63**	−0.11	0.29
26. Not eating because of watching your weight…	0.10	**0.69**	0.02
27. Eat more than usual when see others eating…	0.20	0.08	**0.53**
28. Desire to eat when bored…	**0.74**	0.01	0.15
29. Try not to eat in evening because watching weight…	0.15	**0.69**	0.01
30. Desire to eat when frightened…	**0.75**	−0.03	0.12
31. Take into account weight when eat…	−0.04	**0.89**	−0.08
32. Desire to eat when disappointed…	**0.94**	0.01	−0.02
33. Eating when preparing meal…	0.09	−0.01	**0.45**

	Emot	Restr	Exter
Emot			
Restr	0.38		
Exter	0.59	0.17	

Notes: ESEM, exploratory structural equation modelling; Emot, *Emotional Eating*; Restr, *Restraint*; Exter, *External Eating*. Item with asterisk corresponds to reverse-keyed item. Items with italicized wording indicate problematic items. Shaded cells indicate the factor where the item theoretically belongs. Bold loadings indicate loadings over 0.30.

[†]Exact item wording cannot be shown because of copyright restrictions.

Score reliabilities

The estimates of score reliabilities were computed with the omega coefficient: (i) for *Emotional Eating*, .947 (.954 when item 3 is not considered); (ii) for *Restraint*, .921; and (iii) for *External Eating*, .799 (.818 when item 21 was discarded). These values can be considered adequate for most purposes. For comparability with the results from previous studies, score reliabilities were also computed with Cronbach's alpha. Following the same order, the alphas were .958, .925 and .810.

Table 4 Standardized model results in the MIMIC model (M23)

	Emot	Restr	Exter
Sex[†]	*0.248*	*0.306*	−0.041
BMI	*0.274*	*0.231*	*0.130*
Age	*−0.218*	*0.137*	*−0.308*
Education[‡] (Lower–middle)	−0.074	*−0.129*	−0.063
Education (Higher–middle)	−0.002	*0.084*	−0.011
R^2	0.182	0.194	0.105

Note: Italicized values correspond to statistically significant effects (no correction of *p* values because of multiple comparisons).

[†]*Sex* was coded as 0 = *Men* and 1 = *Women*.

[‡]The effect of level of education was introduced in the analysis with two dummy variables and the middle level of education as the reference category.

Discussion

This study had three goals: (i) to analyse the internal structure of the DEBQ; (ii) to evaluate its measurement invariance; and (iii) to compare the latent means and correlations of some relevant variables. All those analyses were carried out for the first time with a large and nearly representative sample. With respect to the internal structure, our overall results are in line with previous studies (Bozan et al., 2011; Cebolla et al., 2014; Dakanalis et al., 2013; Lluch et al., 1996; van Strien et al., 1986; Wardle, 1987). The DEBQ comprises three related factors. Items about diffuse emotions required correlated uniquenesses, a need that can be undetected when data are analysed using EFA. Given the high redundancy in item wording for some pairs of items, it is not surprising that some additional correlated uniquenesses had to be incorporated, based on modification indices (as in Cebolla et al., 2014). These cross-loadings were low, in general, with the exception of item 3.

We found two items that could be considered problematic: item 3, as a result of nontarget loadings that were almost as high as its target loadings, and item 21, with very low loadings in all the factors. The decisions that should be made about these items depend on the intended uses of the DEBQ. If the instrument is to be used for research, and the data are to be analysed with structural equation modelling, then these items are not problematic, as we have shown that ESEM can fit the data, and the fit of models with or without those items are almost identical. If the DEBQ is to be used for clinical or research analysis carried out with summed scores, our advice would be to delete both items, as the allocation of item 3 to a single dimension is doubtful, and item 21 basically just adds noise. As noted before, when both items are deleted, the reliabilities increase. Having said this, correlations between *Emotional Eating* scores when item 3 was or was not considered (r = .998), and correlations between *External Eating* scores when item 21 was or was not considered (r = .983), were almost perfect, so there is no practical problem with using these questionable items. The final decision should consider whether comparability with other studies using the 33 items of the DEBQ is required. It is important to note that the problems with item 21 (the only reverse-keyed item in the Dutch version) have not been found in versions were this item is direct. A new Dutch version changing the wording of the item is now being considered. The problem with item 3 was

also reported in the Spanish version (Cebolla et al., 2014). The cross-loading of item 28 found in that version was not replicated here.

We were able to establish the measurement invariance of the DEBQ with respect to sex, BMI category, age category and level of education. After evidence of measurement invariance was obtained, replicating and extending the results from Dakanalis et al. (2013), we analysed the relation of the different variables with the factors. The pattern of results followed the expectations derived from theory and previous results (Koenders & van Strien, 2011; Song, Lee, Sung, & Yang, 2013; Sung, Lee, & Song, 2009; van Strien & Koenders, 2012). Most importantly: (i) women and people with higher BMIs had higher scores on *Emotional Eating* and *Restraint*; (ii) *Emotional* and *External Eating* decreased with age, whereas *Restraint* tended to increase; and (iii) the relation between education level and eating styles, although statistically significant in some cases, was negligible.

The significant associations of some of the scales with sex and age imply that norms should be segregated by sex and specific age groups. The finding that emotional eating and dietary restraint were both positively associated with BMI is in line with, respectively, psychosomatic theory and theory of restraint eating (refer to van Strien et al., 1986).

The absence of an association of external eating with BMI is, though not in with externality theory, in line with a body of research summarized in van Strien et al. (2009). The finding that BMI was associated with emotional eating and not with external eating also adds to the evidence showing that emotional and external eating, in spite that they often co-occur, refer to separate constructs (van Strien, Schippers, & Cox, 1995; Ouwens, van Strien, & Van Leeuwe, 2009.

The required reliability of the scores depends on their intended uses (Nunnally & Bernstein, 1994). When scores are to be used for clinical assessment, and individual scores will be interpreted, reliabilities over 0.80 or 0.90 are needed. Reliabilities for *Emotional Eating* and *Restraint* satisfied this cut-off point, whilst *External Eating* did not. The scores from the three scales satisfied the requirement for basic research based on summed scores, where no individual scores are interpreted. The high reliability of *Emotional Eating* and *Restraint* when the main use of the scales is research suggests the possibility of developing a brief version of the DEBQ, whilst still maintaining acceptable levels of accuracy.

Several limitations of this research should be noted. Our results, based on self-reported information, are susceptible to some biases:

(i) for reasons of social desirability, the participants may have disavowed items on emotional and external eating or over-reported dietary restraint; and (ii) the participants may have underestimated their body weight and overestimated their body length, although studies have found a high correlation between self-reported body measures and real measures (e.g. McAdams, Van Dam, & Hu, 2007). It should be noted that item 21 is the only reverse-keyed item in the original DEBQ, and that this item was not reversed in the official English version (van Strien, 2002). We cannot generalize the results about item 21 to these versions. Reverse-keyed items usually present problematic loadings when the internal structure of the scales is assessed (Weijters, Baumgartner, & Schillewaert, 2013). Though our sample is regarded as nearly representative, as it closely follows the distribution in the Dutch population for region, level of education, ethnicity and overweight, there was an overrepresentation of the women.

Amongst the strengths of this paper, we can point to the characteristics of the analysed sample: (i) The sample size of the present sample is 72% of the sum of all the previous samples used in validation studies of the DEBQ; and (ii) the sample can be regarded as nearly representatie of the general population of The Netherlands. With the presented analysis of the DEBQ scores, we have been able to detect some minor problems in its internal structure, which indicate necessary refinements for future versions of the instrument. These problems are not of special relevance in common uses. We have shown that the DEBQ shows adequate measurement invariance, and that the expected pattern of relations with relevant variables was found. Reliability estimates offer some possibilities for further changes, such as short versions. Merging all the information obtained from the different analyses, these results highlight the adequate psychometric properties of the DEBQ.

Ethics statement

Acknowledgements

REFERENCES

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438.

Bozan, N., Bas, M., & Asci, F. H. (2011). Psychometric properties of Turkish version of Dutch Eating Behaviour Questionnaire (DEBQ). A preliminary results. *Appetite, 56*, 564–566.

Bruch, H. (1973). Eating disorders: Obesity, anorexia nervosa, and the person within. New York: Basic Books.

Cebolla, A., Barrada, J. R., van Strien, T., Oliver, E., & Baños, R. (2014). Validation of the Dutch Eating Behavior Questionnaire (DEBQ) in a sample of Spanish women. *Appetite, 73*, 58–64.

COTAN. (2013). Beoordeling Nederlandse Vragenlijst voor Eetgedrag, NVE, 2013 [Review Dutch Eating Behaviour Questionnaire (DEBQ,

2013)]. Retrieved from http://www.cotandocumentatie.nl/test_details.php?id=848

Dakanalis, A., Zanetti, M. A., Clerici, M., Madeddu, F., Riva, G., & Caccialanza, R. (2013). Italian version of the Dutch Eating Behavior Questionnaire. Psychometric proprieties and measurement invariance across sex, BMI-status and age. *Appetite, 71*, 187–195.

Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure, and results. *International Journal of Testing, 10*, 295–317.

Grimm, E. R., & Steinle, N. I. (2011). Genetics of eating behavior. Established and emerging concepts. *Nutrition Reviews, 69*, 52–60.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Koenders, P. G., & van Strien, T. (2011). Emotional eating, rather than lifestyle behavior, drives weight gain in a prospective study in 1562 employees. *Journal of Occupational & Environmental Medicine, 53*, 1287–1293.

Lluch, A., Kahn, J. P., Stricker-Krongrad, A., Ziegler, O., Drouin, P., & Méjean, L. (1996). Internal validation of a French version of the Dutch Eating Behaviour Questionnaire. *European Psychiatry, 11*, 198–203.

McAdams, M. A., Van Dam, R. M., & Hu, F. B. (2007). Comparison of self-reported and measured BMI as correlates of disease markers in US adults. *Obesity, 15*, 188–196.7.

McDonald, R. P. (1999). Test theory: A unified treatment. Mahwah, NJ: Erlbaum.

Muthén, L. K., & Muthén, B. O. (1998–2012). Mplus user's guide (Seventh ed.,,). Los Angeles, CA: Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.,). New York: McGraw-Hill.

Ouwens, M. A., van Strien, T., & Van Leeuwe, J. F. J. (2009). Possible pathways between depression, emotional and external eating: A structural equation model. *Appetite*, *53*, 245–248.

Polivy, J., & Herman, C. P. (1985). Dieting and binging: A causal analysis. *American Psychologist*, *40*, 193–201.

Schachter, S. (1971). Some extraordinary facts about obese humans and rats. *American Psychologist*, *26*, 129–144.

Song, Y. M., Lee, K., Sung, J., & Yang, Y. J. (2013). Changes in eating behaviors and body weight in Koreans: The Healthy Twin Study. *Nutrition*, *29*, 66–70.

Sung, J., Lee, K., & Song, Y. (2009). Relationship of eating behaviour to long-term weight change and body mass index. The Healthy Twin Study. *Eating and Weight Disorders*, *14*, e98–e105.

van Strien, T. (2002). DEBQ. Dutch Eating Behavior Questionnaire. Thames Valley Test Company: Bury St Edmunds.

van Strien, T., Frijters, J. E. R., Bergers, G. P. A., & Defares, P. B. (1986). The Dutch Eating Behavior Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. *International Journal of Eating Disorders*, *5*, 295–315.

van Strien, T., Herman, C. P., & Verheijden, M. W. (2009). Eating style, overeating, and overweight in a representative Dutch sample. Does external eating play a role? *Appetite*, *52*, 380–387.

van Strien, T., Herman, C. P., & Verheijden, M. W. (2012). Eating style, overeating and weight gain. A prospective 2-year follow-up study in a representative Dutch sample. *Appetite*, *59*, 782–789.

van Strien, T., & Koenders, P. G. (2012). How do life style factors relate to general health and overweight? *Appetite*, *58*, 265–270.

van Strien, T., Schippers, G. M., & Cox, W. M. (1995). On the relationship between emotional and external eating behaviour. *Addictive Behaviors*, *20*, 585–594.

Wardle, J. (1987). Eating style: A validation study of the Dutch Eating Behaviour Questionnaire in normal subjects and women with eating disorders. *Journal of Psychosomatic Research*, *31*, 161–169.

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*, 320–334.