Application of the Double Monotonicity Model to Polytomous Items

Scalability of the Beck Depression Items on Subjects with Eating Disorders

Teresa Rivas, Rosa Bersabé, and Carmen Berrocal
Facultad de Psicología, Universidad de Málaga, Spain

Abstract. This paper investigates the item scalability of the Beck Depression Inventory (BDI) in 252 subjects; 126 with and 126 without eating disorders. To do so, an order was established regarding the BDI items according to the clinical characteristics of the subjects with eating disorders. The nonparametric Item Response Theory (NIRT) model was applied to evaluate Monotone Homogeneity and Double Monotonicity of items, as well as the reliability of the scale in both groups. The results show that the order of the items is satisfied in the group with eating disorders, but not in the control group. Therefore, the results obtained allow the ordering of depression scores of subjects with eating disorders according to their clinical characteristics. This order is not valid for the depression scores of subjects who did not have eating disorders. It should be noted that the application of the Double Monotonicity model to polytomous items provides new and relevant information when compared to the data provided by the Classical Test Model. In addition, it is very useful for other items and subjects having certain characteristics.

Keywords: Nonparametric item response theory, polytomous items, depression, eating disorders

Taken as a whole, research on comorbidity indicates a substantial degree of psychological disturbance associated with eating disorders (cf. Halmi et al., 1991). In terms of specific comorbidity, research indicates a high prevalence of affective disorders, especially major depression. Available data from descriptive studies (using structured interviews or psychological tests, as well as family history and antidepressant-medication studies) suggest that depression is a common and significant problem for eating disorder patients (Edelstein & Yager, 1992; Hoffman & Halmi, 1993; Swift, Andrew, & Barklage, 1986).

In fact, because of its association with depression, some authors have hypothesized that eating disorders are variants of affective illness, that is, forms of depression with similar symptoms and the same biological basis as affective disorders (cf. Lee, Rush, & Mitchell, 1985; Pope & Hudson, 1985a, 1985b). However, close examination suggests these arguments are not strong, and that while there is an association between these disorders – in terms of risk and course – there are also features that distinguish them (Cooper, 1995; Hinz & Williamson, 1987).

While the clinical characteristics of eating disorders have been described in great detail, there have been no systematic phenomenological studies of the depressive features of these disorders. This study is concerned with the character of depressive symptoms in these patients. Consistent with the hypothesis that depressive symptoms are secondary to the eating disorder itself, the present study tested whether depressive symptoms of subjects with eating disorders can be ordered according to the main clinical features of these patients.

In order to assess depressive symptomatology, the Beck Depression Inventory (BDI; Beck, Rush, Shaw, & Emery, 1979; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) was used in this study. The BDI includes 21 items that assess the presence and severity of cognitive ideation (feeling of sadness, pessimism, feeling like a failure, dissatisfaction, feeling guilty, feeling punished, disappointed with oneself, self-critical, suicidal ideation, crying, irritability, and feeling unattractive); somatic manifestations of depressive symptomatology (sleep disturbance, feeling of lack of energy, worry about phys-
ical health, loss of interest in sex, and appetite and weight loss); and motivational ones (loss of interest in people, difficulty in making everyday decisions, and retarded initiation of voluntary responses).

Because anorexia nervosa and bulimia nervosa are mainly characterized by an excessive concern about body image and weight (Fairburn & Garner, 1988; Garfinkel, 1992), it was expected that the main depressive symptoms would be related to self-accusations, self-image, or self-dislike. Likewise, irritability and feeling sad should be common in patients because they are invariably distressed by their efforts to control their weight and eating. Similarly, because the physical consequences of the various abnormal eating-related behaviors are common in these patients (Kaplan & Garfinkel, 1993), it was expected that retarded initiation of voluntary responses, insomnia, or fatigue would be among the main depressive symptoms.

On the other hand, despite the medical complications of eating disorders, patients present disturbances in the way in which their body weight or shape is experienced, or they deny the seriousness of their current low body weight, specific problem behaviors, and the physical consequences (Fairburn & Garner, 1988; Garfinkel, 1992). Therefore, it was expected that depressive symptoms such as hypochondria, libido loss, anorexia, or weight loss would be at the bottom of the scale. Finally, because dieting and other eating behaviors—once commenced—may lead to feelings of success and “being in control,” and because of the social dependence that characterizes these patients (cf. Slade, 1982), it was expected that withdrawal, feeling punished, feeling like a failure, or pessimism would also be among the less important symptoms.

It was supposed that items of the BDI would be ordered (from least to most important) on subjects with eating disorders as follows: Hypochondria (Item 20), libido loss (21), withdrawal (12), self-punishment (6), anorexia (18), weight loss (19), failure (3), pessimism (2), crying (10), suicidal feelings (9), guilt (5), indecision (13), dissatisfaction (4), fatigue (17), insomnia (16), retarded responses (15), sadness (1), irritability (11), self-dislike (7), self-accusations (8), self-image (14). On the other hand, it was assumed that the order of items established for a group (with eating disorder) would not necessarily be valid for another with different characteristics (Control group).

BDI is an instrument widely used to assess depression in subjects with other nonaffective primary disorders (e.g., eating disorders). Therefore, it would be interesting to test whether or not it is possible to consider a specified order in BDI items to assess depression in subjects with eating disorders. This would imply that the BDI is an accumulative scale, i.e., an ordering on the items can be established, and it should be taken into consideration when obtaining the total BDI score. In this respect, Rivas, Bersabé, and Berrocal (2001) show preliminary results of this study.

This paper presents a study of the scalability of BDI polytomous response items (0–3) from responses given by subjects with eating disorders (Clinical group) and others without eating disorders (Control group). In each of the groups, separately, item scalability is analyzed by the Double Monotonicity (DM) model for polytomous items. The fit of this Nonparametric Item Response Theory (NIRT) model gives indices of evaluation of the Monotone Homogeneity (MH) model and the DM model as well as for the reliability of the set of items. By fitting this new model, detailed information is obtained about the scalability of items and subjects of each of the groups considered (with and without eating disorders).

**The DM Model for Polytomous Items**

Within the framework of the NIRT, Mokken (1971, 1997) defines MH and DM models for dichotomous items. The DM model is one of the most attractive models available in psychometrics. This is because a unidimensional set of items verifying some nonparametric assumptions satisfies the conditions of the DM model. That is to say, a set of items that satisfies unidimensionality and local independence, and whose Item Response Functions (IRFs) are nondecreasing monotone, verifies the assumptions of MH model. If, in addition to the foregoing conditions, IRFs do not intersect, the set of items also verifies the assumptions of DM model. In a unidimensional scale, this latter model permits the ordering of items independently from the subjects, and vice versa.


Checking whether a set of items satisfies the assumptions of the DM model is laborious. This means that in practice, the fit of model was not viable until Molenaar, Debets, Sijtsma, and Hemker (1994) developed and published software (MSP 3.0) to check if a set of items satisfies DM. Later, Molenaar, Sijtsma, van Schuur, and Mokken (2000) developed a new improved version (MSP5 for Windows) adding new indices of model goodness-of-fit. Previous references together with Hemker, Sijtsma, and Molenaar (1995), Hemker, Sijtsma, Molenaar, and Junker (1997), Sijtsma (1998), Molenaar and Sijtsma (2000), Sijtsma (2001), Sijtsma and Molenaar (2002) give an overall idea about these models, the evaluation indices, and their applications. Owing to the
recent development of these models and the program to fit model data, applications of these models are scarce.

Molenaar et al. (1994) and Molenaar and Sijtsma (2000) set out the indices that evaluate whether a set of items satisfies the MH or DM and suggest additional reading regarding the detailed study of some evaluation indices.

To evaluate if IRFs are nondecreasing monotone or satisfy single monotonicity there are:

- Scalability coefficients based on the analysis of each item \(H_i\), item pairs \(H_{ij}\) or a set of items \(H\) (Mokken, 1971).
- Indices for each item obtained in the entire group.
- Indices for each item obtained in restscore groups. These groups are defined on the scores of the remaining items. Given a increasing restscore \(r\), the proportion of positive responses must be monotonically non-decreasing in \(r\) (Rosenbaum, 1984; Molenaar et al., 1994, p. 9).
- Diagnostic Value Crit, which summarize information obtained through the checking of the single monotonicity via the entire group, enabling the identification of items least fitted by the MH model. (Molenaar & Sijtsma, 2000, pp. 49, 74).

Indices that evaluate the nonintersection of IRFs are usually based on the analysis of item pairs. They check the nonintersection via:

- P-matrices by visual inspection (Mokken, 1971) and by a count of the violations. Given an ordering on the items, a MH set of items is DM when columns and rows in \(P_1\) matrix are monotonically nondecreasing, and columns and rows in \(P_0\) matrix are monotonically nonincreasing. Local deviations from these orders are considered violations of DM.
- Indices for item pairs obtained in restscore groups. These restscore groups are determined on the remaining items. The proportion of positive responses on an item should be smaller than or equal to that of the other item in each restscore group. (Rosenbaum, 1987; Molenaar et al., 1994, p. 10).
- Indices for item pairs obtained in restsplit groups. These groups are determined by distinguishing between counts for the low and high groups that are based on the use of cutpoints. (Molenaar, 1991; Molenaar & Sijtsma, 2000; Sijtsma & Junker, 1996, pp. 88–89).
- Diagnostic Value Crit that summarize information obtained through the checking of the nonintersection via restscore groups and via P-matrices, enabling the identification of items least fitted by DM model.
- \(H^2_T\) and \(H^2_S\) coefficients. \(H^2_T\) analyses if a set of items have intersecting IRF. \(H^2_S\) for the total set of items based on the transposed data matrix and \(H^2_T\) coefficients on the level of the individuals are determined. Rules of thumb for their interpretation were based on results from a study using simulated data. (Sijtsma & Meijer, 1992).

- Graphical representations of IRFs can be shown and it can be seen if pairs of items intersect.

These indices make it possible to check whether or not IRFs are monotone and intersect. If a set of items satisfies the assumptions of the MH model, the IRFs are non-decreasing monotone functions. Then an order on the subjects can be considered. In addition, if pairs of IRFs do not intersect, the set of items MH is DM, so items can also be ordered. Only if the set of items satisfies the conditions of the DM model, can the reliability of a set of items be interpreted.

To evaluate the fit of MH and DM models, the MSP5 program of Molenaar, Sijtsma, van Schuur, and Mokken (2000) was used. One of two procedures can be chosen to fit model data. The Search procedure, which explores the underlying order on the items, and the Test procedure, which proves whether a preestablished ordering on the items can be accepted or not.

The Test procedure was used to ascertain whether the items could be ordered in the clinical and control group, according to the order previously established.

The following steps then guided the procedure to obtain the scalability and reliability of the items of the BDI in both groups. In the interpretation of results, the following boundaries given by Mokken (1971) and Molenaar and Sijtsma (2000) were taken into account.

1) Evaluating whether IRFs are nondecreasing monotone is done by:

1.1 Scalability coefficients \(H_{ij}, H_i\) and \(H\). Scalability coefficients must satisfy the bounds \(H_{ij} > 0, H_i \geq 0.30, \) and \(H \geq 0.30\) (Mokken, 1971, p. 184).

1.2 Checking the monotonicity in the entire group. This evaluation is made taking into account the size of violations \(#vi/#ap\) of monotonicity. The violations are evaluated by means of the differences between two such proportions when their ordering is not in agreement with the requirements of the model (Molenaar et al., 1994, p. 45). It is also admitted that the minimum size of violations is, by default, 0.03 (Molenaar & Sijtsma, 2000, p. 66).

1.3 Diagnostic Crit Value

A Diagnostic Crit Value which exceeds 80 casts serious doubt on the validity of the model for this item. If it is lower than 40, the violations reported may well be ascribed to sampling variation. Between 40 and 80 a decision may depend on a fur-
Other consideration of the pros and cons. In addition, the program indicates with an asterisk (*) the largest Crit Value (Molenaar & Sijstma, 2000, pp. 49, 74).

2) Nonintersection is investigated by a check of violations via:

2.1 A global study to summarize the information of P-matrices.

2.2 Restscore groups. In this step it is also admitted that the minimum number of subjects that must have a restscore group is, by default, N/5. As in 1.2, in 2.1, and 2.2, the minimum size of violations used in this study is 0.03.

2.3 Restsplit groups. #sig low and #sig high give the number of significant violations (Molenaar & Sijstma, 2000, p. 81).

2.4 Diagnostic Crit Values

Interpretation is similar to that given in 1.3 above.

Method

Participants

The sample consisted of 252 subjects: 126 cases attending treatment for an eating disorder in different clinical centers, and 126 controls from the general population without eating disorders. In both groups, there were 113 (89.7%) females and 13 (10.3%) males, aged between 12 and 45 (\( M = 19.23, SD = 5.35 \) in cases; and \( M = 19.60, SD = 5.98 \) in controls). Cases with an eating disorder were in different phases of treatment: 91 (72.2%) in assessment, 32 (25.4%) in cognitive-behavioral treatment, and 3 (2.4%) in follow-up treatment.

205 subjects responded to all items of the BDI (92 cases and 113 controls). The subjects’ classification by BDI scores (Beck, Steer & Garbin, 1988) can be seen in Table 1.

<table>
<thead>
<tr>
<th>BDI</th>
<th>Clinical group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (( X_i \leq 9 ))</td>
<td>29</td>
<td>81</td>
</tr>
<tr>
<td>Slight (10 ( \leq X_i \leq 18 ))</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>Moderate (19 ( \leq X_i \leq 29 ))</td>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td>Severe (30 ( \leq X_i \leq 63 ))</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>92</td>
<td>113</td>
</tr>
</tbody>
</table>

Table 1. Subjects’ classification by BDI scores.

Materials

The clinical diagnosis of eating disorder was established using the Spanish version of the Questionnaire for Eating Disorder Diagnoses (Q-EDD) by Mintz, O’Halloran, Mulholland, and Schneider (1997) (Rivas, Bersabé, & Castro, 2001). The Q-EDD consists of 50 items aimed at measuring DSM -IV criteria for eating disorders (APA, 1994). The format of the responses changes from one item to another; some follow a dichotomous model (yes/no), others offer 4, 5, or 6 alternative responses and, on a few occasions, there is an open question. The 50-question Q-EDD operationalizes eating disorder criteria of the DSM-IV. Diagnoses are generated by a scoring manual that consists of flowchart decision rules. Subjects with an eating disorder are classified in these categories: anorexia nervosa, bulimia nervosa, and eating disorders not otherwise specified (EDNOS). Subjects without an eating disorder are classified as symptomatic or asymptomatic. Q-EDD diagnoses in clinical and control groups are given in Table 2.

<table>
<thead>
<tr>
<th>Q-EDD diagnosis</th>
<th>Clinical group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anorexia nervosa</td>
<td>8</td>
<td>6.3</td>
</tr>
<tr>
<td>Bulimia nervosa</td>
<td>20</td>
<td>15.9</td>
</tr>
<tr>
<td>EDNOS</td>
<td>72</td>
<td>57.1</td>
</tr>
<tr>
<td>Symptomatic</td>
<td>14</td>
<td>11.1</td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>7</td>
<td>5.5</td>
</tr>
<tr>
<td>Missing</td>
<td>5</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Table 2. Diagnoses in clinical and control groups.

Depression was evaluated by BDI (Beck et al., 1961, 1979). The BDI is the most commonly used self-report measure of this symptomatology. Although it is not necessarily indicative of the full clinical syndrome of depression, it is a reliable and valid measure of many depressive symptoms (Canals, Blade, Carbajo, & Domeche, 2001). Beck et al. (1961) reported a split-half reliability coefficient of .93 and high correlations between clinical ratings of patients and their scores on the BDI (Beck, 1967). Previous research has also established the BDI’s reliability and validity in normal samples (Beck et al., 1988). Likewise, two studies have been conducted in order to investigate the factor structure of the BDI in subjects with eating disorders (Pulos, 1996; Welch, Hall, & Walkey, 1990). The results of both studies show that the BDI appears to assess a unidimensional construct (depression) in patients with eating disorders.

Procedure

The subjects in the clinical group (with eating disorders) were selected from a wider sample of subjects receiving assistance in several clinical centers. The professionals working in these centers (psychologists, endocrinologists, pediatricians, psychiatrists, etc.) informed the research group about each case of a possible eating disor-
The research group traveled to the different centers in order to provide the instruments already mentioned. Both instruments were self-administered. According to the results obtained with the Q-EDD, we selected the subjects who satisfied all the criteria for a diagnosis of any of the eating disorders considered in the DSM-IV. In addition, in each case we asked the patient to ask a friend without an eating disorder (as far as they knew), but of the same sex and a similar age, to come to the center to complete the same questionnaires. When the subject could not bring anyone, the research group arranged for a person with similar characteristics and belonging to the general population to complete the questionnaires. The evaluation session always began by measuring the subject’s height and weight.

**Results**

Before checking the DM model, it is necessary to test the assumption of unidimensionality of items. The results of evaluation of Unidimensionality, MH, DM, and the reliability of the scale are shown, using the procedure outlined above.

**Unidimensionality**

Unidimensionality is analyzed by Principal Component Analysis. The first three eigenvalues are given in Table 3.

<table>
<thead>
<tr>
<th>Eigen</th>
<th>Clinical group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.4727</td>
<td>5.3185</td>
</tr>
<tr>
<td>2</td>
<td>1.3855</td>
<td>1.7044</td>
</tr>
<tr>
<td>3</td>
<td>1.2169</td>
<td>1.5906</td>
</tr>
<tr>
<td>UNIDI</td>
<td>47.97</td>
<td>31.76</td>
</tr>
</tbody>
</table>

Based on index $\text{UNIDI} = (\lambda_1 - \lambda_2)/\lambda_3 > 5$ (Martínez Arias, R., 1995, p. 297), the items clearly satisfy unidimensionality in all the groups. The loadings of the items in the factor matrix are greater than 0.32 in each group, except for Item 20 whose loadings are lower than 0.30 (see Table 4). The communality of Item 20 is very small ($< 0.1$) in all the groups. The percentage of variance accounted for by the factor is 45.1 and 25.3 in each group, respectively (see Table 4).

**Evaluation of the MH Model**

Results of evaluating whether IRFs are nondecreasing monotone are shown in Table 5.

1) **Clinical group.** In Column 1 of Table 5 it can be seen that the ordering preestablished on the items (Column 1) is the same as if the ordering had been established by the means of the items (Column 2), except for Item 20. In Column 3, $H_i$ coefficients for each item are shown. They assume values from 0.33 (Item 19) to 0.57 (Item 7). All the items have been selected for the MH model because the scalability coefficients satisfy the bounds $H_i > 0$, $H_i \geq 0.30$ (except for Item 20 with value $H_i = 0.16$), and $H_i = 0.45 \geq 0.30$.

The size of violations in the entire group varies from 0.07 for item 20 to 0.31 for Item 2 (Column 4). The greatest Crit values (Column 5) are 33 (Item 20 and Item 18), 39 (Item 8), 67 (Item 2), and 71 (Item 11). Although there are violations greater than 0.03, their Crit values are less than 80. Therefore, all the IRFs of the BDI are nondecreasing in this group, except for Item 20.

2) **Control group.** In Table 5 it can be seen that the ordering preestablished on the items (Column 1) is not the same as if the ordering had been established by the means of the items in this group (Column 2), unlike the Clinical group. All the items except for 1, 5, 8, and 10 have $H_i < 0.30$ values, the lowest of which is $H_i = 0.06$ corresponding to Item 20 (Column 7 in Table 5).

The size of violations is zero for all items in the entire group, except for Item 3 and Item 20 (Column 8), whose...
Crit values are 73 and 189 (Column 9), respectively, this latter value being significant.

In summary, \( H_i \) values are greater than 0.30 for all items in the Clinical group, except for Item 20. However, a large number of items have \( H_i < 0.30 \) in the Control group. The lack of homogeneity between Item 20 and the rest of the items (\( H_i = 0.06 \)) is particularly notable in the Control group. In regard to the significant Diagnostic Crit value, Item 20 in the Control group stands out because of its Crit value of 189.

Thus, MH is satisfied in the Clinical group. Better indices for \( H_i \) and \( H \) are required for the set of items to satisfy MH in the Control group.

### Evaluation of the DM Model

Nonintersection of the item pair Response Functions is investigated in this section. The results are shown in Table 6.

1) **Clinical group.** In Column 2 of Table 6, checking nonintersection via P-matrix, all the violations are \( \leq 0.03 \) except for Item 19 (\( \#vi/#ap = 0.04 \)) and Item 20 (\( \#vi/#ac = 0.08 \)) whose significant Crit values are 133 and 230, respectively (Column 3). In addition there are Crit values greater than 80 in other items (Column 3). In restscore groups, (Columns 4 and 5 in Table 6) violations are seen between 0.05 (Item 15) and 0.19 (Item 20) (Column 4). The size of violation for Item 20 is 0.19, whose Crit value is 160 and, therefore, it is the largest critical value for this item set (Column 5). Other items whose Crit values are greater than 80 can be seen in Column 5.

In restsplit groups, (Columns 6 and 7 in Table 6) it can be seen that items which previously had Crit values greater than 80 now only have a small number of significant violations:

- There are a few important violations – given in brackets – at the lower end of the scale (low): Item 1 (1); Items 20, 18, 19, 8 (2); and Item 3 (3).
- There are a few important violations – given in brackets – at the higher end of the scale (high): Items 9, 5, 17, 7, and 8 (1); Items 1 and 4 (1); and Item 20 (7).

2) **Control group.** All violations from nonintersection via P-matrix are less than 0.03 (Columns 8 in Table 6); the only significant Crit value being 81 for Item 20 (Column 9). In restscore groups, (Columns 10 and 11 in Table 6) the size of violations goes from 0.02 to 0.30; 0.30 being obtained in Item 20 and its Crit value 125 is significant.

In restsplit groups (Columns 12 and 13 in Table 6):

- At the lower end of the scale (low), the number of important violations – given in brackets – is in Item

### Table 5. Means and indices to evaluate MH model assumptions.

<table>
<thead>
<tr>
<th>Items</th>
<th>Clinical group</th>
<th>Entire group</th>
<th>Control group</th>
<th>Entire group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>( H_i )</td>
<td>#vi/#ap</td>
<td>Crit</td>
</tr>
<tr>
<td>20</td>
<td>0.82</td>
<td>0.16*</td>
<td>0.07</td>
<td>33</td>
</tr>
<tr>
<td>21</td>
<td>0.59</td>
<td>0.42</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.64</td>
<td>0.48</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.66</td>
<td>0.45</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.70</td>
<td>0.34</td>
<td>0.12</td>
<td>33</td>
</tr>
<tr>
<td>19</td>
<td>0.72</td>
<td>0.33</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.72</td>
<td>0.52</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.73</td>
<td>0.44</td>
<td>0.31</td>
<td>67</td>
</tr>
<tr>
<td>10</td>
<td>0.78</td>
<td>0.44</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.80</td>
<td>0.56</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.96</td>
<td>0.46</td>
<td>0.09</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>0.90</td>
<td>0.49</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.97</td>
<td>0.50</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.99</td>
<td>0.45</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1.00</td>
<td>0.41</td>
<td>0.13</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>1.07</td>
<td>0.49</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.01</td>
<td>0.53</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.08</td>
<td>0.44</td>
<td>0.23</td>
<td>71</td>
</tr>
<tr>
<td>7</td>
<td>1.05</td>
<td>0.57</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.09</td>
<td>0.56</td>
<td>0.30</td>
<td>39</td>
</tr>
<tr>
<td>14</td>
<td>1.21</td>
<td>0.42</td>
<td>0.06</td>
<td>12</td>
</tr>
</tbody>
</table>

| 0.45 | 0.24* |
At the higher end of the scale (high) there are no important violations (Column 13, Table 6).

In summary, the indices which check the DM give the following:

1) In the Clinical group, Item 20 has the largest Crit value when checking the nonintersection via P-matrix and via restscore groups. In addition, via restsplit groups, Item 20 has the greatest number of significant deviations (7) at the higher end of the scale.

In the Clinical group, the items of the BDI (except for Item 20) satisfy the assumptions of MH and DM. The reliability of the scale is 0.94 in this group. Therefore, the set of items is DM and the scale is reliable in this group.

2) In the Control group, Item 20 has a significant Crit value of 81 when checking nonintersection via P-matrix. Via restscore groups, Item 20 has a significant Crit value. Via restsplit groups, Item 20 has the greatest number (4) of significant deviations at the lower end of the scale.

In the Control group, IRFs do not intersect (except for Item 20), but many items of the BDI do not satisfy the assumptions of MH model. Therefore, it is not possible to establish this ordering on the items, the set of items is not DM and the reliability of the scale cannot be interpreted in this group.

### Descriptive Statistics of Items of the BDI

The SD shows that there is variability in the scale score in both groups (Tab. 7). The frequency distribution of the scale score is shown in Figure 1.

### Discussion

A higher level of depression was found in the Clinical group than in the Control group. In the former, 68.5% of the subjects scored 17 or greater on the BDI, as compared to 28% in the Control group. This finding is consistent with that obtained in clinical and university samples of subjects with eating disorders and confirms that patients...
with eating disorders have a significant degree of depressive morbidity (cf. Dykens & Gerard, 1986; Schlesier-Carter, Hamilton, O’Neil, Lydiard, & Malcolm, 1989). On the other hand, unidimensionality of items is clearly obtained in both groups. This finding also confirms the results of earlier studies conducted in clinical populations with eating disorders, which showed the presence of a large general factor only, and suggests that the items of the BDI recognize a general construct of depressive symptoms rather than a variety of more specific constructs (Pulos, 1996; Welch, Hall, & Walkey, 1990). It should be noted that only Item 20 has loading smaller than 0.30 and a small communality (< 0.01) in these groups.

MH is satisfied except for Item 20 in the Clinical group. This result suggests that the IRF of hypochondriasis (Item 20) is not monotonically nondecreasing in this group. However, MH is not satisfied in the Control group, because a large number of items do not have monotonically nondecreasing IRF. Therefore, in the Control group, items do not satisfy the preestablished order. The IRF do not intersect in any group (except for Item 20).

The scale is reliable in the Clinical group. It is possible to use this order of items in the Clinical group, and also to order the subjects according to this continuum of ordered symptoms of depression. This is not possible in the Control group because the set of items does not satisfy MH in this group.

Regarding the scalability of items in the Clinical group, the hypothesized order has been confirmed. In addition, this order coincides with the order that would have been established from the means of these items (except for item 20) through an exploratory study. Thus, the results of this study suggest that the most important depressive symptoms in eating disorder patients are negative cognitions related to self-accusation, self-image, and self-dislike. As some authors have argued, these features might be secondary to the core of eating disorders, that is, the extreme concern with body shape and weight (Fairburn & Garner, 1988). In this sense, the results are consistent with the assumptions underlying the cognitive-behavioral approach applied to eating disorders. Reduced to its essence, this approach assigns a primary role to weight and shape concerns, holding that eating disorder symptoms are maintained by a characteristic set of beliefs about weight and eating.

Likewise, affective features such as irritability and feeling of sadness, as well as motivational symptoms (reduced initiation of voluntary responses, insomnia, or fatigue), were evident in the pattern of the mainly depressive symptoms. They can arise as a consequence of the medical complications of eating disorders.

On the other hand, people with eating disorders also have symptoms of libido loss, anorexia, weight loss, and a wide range of physical symptomatology. However, and consistent with the initial hypothesis, these symptoms point to the bottom of the scale. The results are consistent with the features of eating disorders. Patients with these disorders may be less likely to report this symptomatology due to some of the disturbances of the eating disorder itself. Thus, patients commonly deny the seriousness of their current low body weight, specific problem behaviors, and physical symptomatology that arise as a consequence of their maladaptive eating behaviors.

In the same way, withdrawal, feeling punished, feeling like a failure, or pessimism were also among the less important symptoms. This pattern may be consistent with Slade’s formulation (1982), among others. In summary, Slade suggests that once dieting and other behav-

Figure 1. Frequency distributions of BDI scale scores in the different groups.

![Figure 1](image-url)
iors to control body weight have commenced, one type of consequence may come into play, which has the effect of strongly reinforcing these behaviors, leading to a downward spiraling effect: feelings of success, feelings of being-in-control, and feelings of self-satisfaction. These feelings might be incompatible with those of failure or pessimism and can be viewed as operating as positive reinforcers.

In general, the results indicate that depressive symptomatology in patients with eating disorders is anything but pessimistic in nature. It is fundamentally related to negative cognitions, and may be largely secondary to the eating disorder itself.

As anticipated, the differences between the Clinical and the Control groups regarding the ordered symptoms reflect the fact that the types of depression in both groups are not the same. As previously noted, a large number of subjects in the Control group fell within the range of normality, while a large number of subjects in the Clinical group fell within the range of depression. Given the great variability of behaviors measured by the BDI in normal samples, it is difficult or nearly impossible to order depressive symptoms in the former. In fact, although the BDI has been argued as having validity in nonclinical populations, the use in normal samples of instruments which were developed in clinical populations has been questioned by some authors (cf. Dohrenwend, Shrut, Egri, & Mendelsohn, 1980). It is possible that, despite the utility of the BDI to assess the level of cognitive, affective, motivational, or somatic features, there is no pattern of these behaviors in nondepressive individuals. The results obtained in this study are one step toward resolving the question put forward by other authors (Hammond, 1995).

It would be of interest to extend this study to other clinical groups presenting depression, with the high likelihood that the order of symptoms would be different to that presented in this study. The advantage of considering a particular order for depression symptoms in each clinical group is that we can then compare scores obtained for exactly the same symptoms. Thus, two subjects who obtain the same score on the BDI will also be similar regarding the symptoms that they scored on. The procedure followed in this paper presents new information regarding the internal validity of the test (BDI) in the group studied, which in turn complements the results of other current reliability and validity studies.

Acknowledgment

This research was funded by the Ministerio de Ciencia y Tecnología, Proyecto ‘Cuestionarios para el Diagnóstico de los Trastornos de la Conducta Alimentaria’ (Ref. No. BSO2001–1945).

References


Address for correspondence

Teresa Rivas, Rosa Bersabé, and Carmen Berrocal
Facultad de Psicología
Universidad de Málaga
Campus de Teatinos
E-29071 Málaga
Spain
Tel. +34 95213-1090
Fax +34 95213-2621
E-mail moya@uma.es