Author: Wardenaar, Klaas Johan
Title: Syndromes versus symptoms : towards validation of a dimensional approach of depression and anxiety
Date: 2012-10-04

Chapter 3:

The Structure and Dimensionality of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in Patients with Depressive Disorders and Healthy Controls

Abstract

Background: The inventory of depressive symptomatology self report (IDS-SR) is a widely used but heterogeneous measure of depression severity. Insight in its factor structure and dimensionality could help to develop more homogeneous IDS-SR subscales. However previous factoranalytical studies have found mixed results. Therefore, the present study tested which factor structure underlies the IDS-SR and, in addition, if the factors can be used as unidimensional subscales. *Methods:* Confirmatory factor analysis (CFA) was done to identify the best fitting factor structure. The study sample consisted of 2600 individuals (mean age 40.5 ± 12.1). We assessed model fit in 4 groups: 957 Major Depressive Disorder (MDD) patients, 450 remitted MDD patients, 570 patients with an anxiety disorder and 623 healthy controls to test the consistency of model fit. Rasch analyses in the full sample were used to evaluate and optimize the unidimensionality and psychometric quality of the factors.

Results: CFA indicated that a 3-factor model fits the IDS-SR data best and is consistent across groups, with a 'mood/cognition' factor, an 'anxiety/ arousal' factor and a 'sleep' factor. In addition, Rasch analyses indicated that the 'mood/cognition' and 'anxiety/arousal' factors could be optimized to be used as unidimensional subscales.

Limitations: The fit of only 4 models was tested, ranging from a 1- to 4-factor model.

Conclusions: The IDS-SR is a heterogeneous instrument with a multifactorial underlying structure. It is possible to measure more homogenous symptomatology with IDS-SR subscales, which could be useful in clinical practice and scientific research.

Previously published as:

Wardenaar, KJ, Van Veen, T, Giltay, EJ, Den Hollander-Gijsman, ME, Penninx, BWJH, Zitman, FG (2010). The structure and dimensionality of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in patients with depressive disorder and healthy controls. *Journal of Affective Disorders* 125: 146-154.

3.1 Introduction

Major Depressive Disorder (MDD) is a heterogeneous disease; patients vary in terms of severity, age of onset, duration, recurrence and symptom profiles (Kendler, 1999). Consequently, the 'depression' label provides limited information about the particular problems experienced by a patient. The description of a patient's condition in terms of symptom dimensions creates possibilities for more specific diagnosis, treatment evaluation and research (Shafer, 2006; Andrews et al., 2007). Therefore, in the present study we evaluated which factor structure underlies the widely used Inventory of Depressive Symptomatology self report (IDS-SR) and whether the identified factors could be used as unidimensional subscales.

The IDS-SR is a self report questionnaire that was introduced by Rush et al (1986; 1996) as a measure of depression severity. The IDS-SR comprises all symptoms of depression, as defined by the DSM-III-R / IV-TR, including melancholic, atypical and anxious symptoms. The questionnaire has been shown to have adequate reliability, acceptable validity, good responsiveness and good discriminative ability (Rush et al., 1996, 2003; Corruble et al., 1999; Trivedi et al., 2004).

The IDS was aimed to measure a unitary construct of depressive symptom severity (Rush et al., 1996; Trivedi et al., 2004). However, different underlying factor structures have been described for the IDS-SR, with solutions of two factors (Bernstein, 2006), three factors (Rush et al., 1996) and four factors (Rush et al., 1986). Although the numbers of factors are different, the models show conceptual overlap: all the multifactorial models make some distinction between a 'depression/mood' and an 'anxiety/somatic' dimension (Bernstein, 2006; Rush et al., 1986; 1996). In addition, the 3-dimensional model contains a 'sleeping problems' factor and the 4-dimensional model contains an 'atypical' and an 'endogenous' factor.

Differences between the presented structures may be attributed to differences in characteristics of the analyzed samples and analytic approaches (Fabrigar et al., 1999). One study (Bernstein, 2006) used principal components analysis (PCA) to explore the underlying structure of the IDS-SR. Two studies used PCA with a predefined number of components to be extracted (Rush et al., 1986, 1996). Yet, confirmatory factor analysis (CFA) would provide valuable information about the appropriateness of hypothesized factor structures. Trivedi et al (2004) used CFA to test the fit of a 1-factor model to a large set of items, including the IDS-SR, and found it to fit well. However, CFA should ideally be used to investigate and compare the fit of several hypothesized structural models (Brown et al., 2006).

A stable factor model for the IDS-SR would have a potential utility in patient care and to address specific research questions. It could help uncover clusters of symptoms that are systematically related to each other, which could potentially be used as IDS-SR subscales. However, before a set of items is used as an additive interval scale, it should be determined if it is really unidimensional. This can be investigated with Rasch modelling methods. The Rasch model is an Item Response Theory (IRT) model that models the probability of endorsement of each item in an instrument as a function of its location on the underlying symptom-severity dimension. If all items fit adequately to the Rasch model, this indicates that the items are ordered along one dimension and that the added up raw, ordinal item-responses can be interpreted as a true interval scale (Wright & Masters, 1982). In addition, Rasch analyses can be used to investigate the discriminative ability of a measure, and whether items function consistently across different person characteristics. Importantly, if the factor structure and the fit to the Rasch model are consistent across healthy subjects and patients, this would indicate that each subscale measures the same underlying severity dimension, irrespective of an individual's categorical diagnosis.

We aimed to find the best-fitting, most consistent factor structure for the IDS-SR and to investigate the usability of the factors as subscales. To this end, we tested and compared the fit of four factor models from the literature using CFA in patients with current MDD (n=957). Next, we investigated the consistency of the best-fitting model across patients with a remitted MDD (n=450), patients with a lifetime anxiety disorder (n=570) and healthy control subjects (n=623). Rasch analyses were conducted to evaluate unidimensionality of the factors and to investigate whether the added up, raw responses on each factor could be used as subscales with sufficient discriminative ability and stability across different person characteristics. We conducted these analyses on data from the Netherlands Study of Depression and Anxiety (NESDA; N = 2981).

3.2 Methods

Sample and procedures

Participants came from the NESDA study, a large scale longitudinal study conducted among 2981 adult subjects (mean age 41.9, age range: 18-65; 1002 men and 1979 women) (Penninx et al., 2008). The NESDA sample consists of 2329 subjects with a lifetime diagnosis of depressive or anxiety disorder and 652 subjects without a lifetime psychiatric diagnosis. These were recruited from three different settings: community, primary care and mental health care organizations. All participants were interviewed and assessed during a visit to a research location.

For the CFA analyses, subjects that completed the IDS-SR without missing values (n=2600) were included in one of 4 non-overlapping groups. Group 1 consisted of all subjects with a current MDD diagnosis (within last 6 months; with or without MDD in the past; n=957). Group 2 consisted of subjects with MDD in remission during the past 6 months; n=450). Group 3 (n=570) consisted of patients with a lifetime anxiety disorder and no lifetime depression. Group 4 (n=623) consisted of all mentally healthy control subjects. The protocol of the NESDA study was approved centrally by the Ethical Review Board of the Leiden University Medical Center and by local review boards of participating centres. All subjects signed informed consent.

Instruments

All participants completed the Dutch translation of the IDS-SR (Rush et al., 1996) that consists of 30 equally weighed items, rated on a four-point scale (range 0-3). 28 of the 30 items are summed to a standard total score, ranging from 0 to 84 (as only appetite and weight increase or decrease is scored). For the analyses of the 1, 2 and 3-factor models and the Rasch analyses, items 11 and 12 and items 13 and 14 were rescored into a single 'change of appetite' (item 11/12) variable and a 'weight change' variable (item 13/14). Only for the analysis of the 4-factor model, items 11, 12, 13 and 14 were treated as separate variables. The Composite International Diagnostic Interview (CIDI, WHO version 2.1) was used to assess the DSM-IV criteria for depressive disorders (MDD and dysthymia) and anxiety disorders (panic disorder, social phobia, generalized anxiety disorder and agoraphobia).

Statistical analyses

Principal Component Analysis

To gain some preliminary insight into the number of components that could be expected to underlie the IDS-SR, an initial PCA was conducted in all subjects that completed the IDS-SR (n=2600). Parallel analysis was used for factor extraction. With this method, the eigenvalues generated with PCA in the real data are compared with the (95th percentile of) eigenvalues that are generated in 1000 random datasets with the same number of variables and observations. Only the components are retained for which the eigenvalues in the real data exceed the randomly generated eigenvalues, and thus are higher than expected by chance. This method has been shown to be superior to traditionally used extraction techniques like Kaiser's criterion (O'Connor, 2000). PROMAX was used for oblique component rotation. The analyses were conducted with SPSS (version 16).

Confirmatory factor Analysis (CFA)

To investigate and compare the fit of four factor-models from the literature, CFA was conducted in current MDD patients (group 1), because all four hypothesized models were developed in comparable samples of depressed (out)patients. To test model-fit with CFA, the models were first defined based on the (PCA) factor loadings reported in each publication (Rush et al., 1986, 1996; Trivedi et al., 2004; Bernstein, 2006). For each item it was evaluated on which factor it had its highest (primary) loading in the PCA results. In the CFA model, this primary loading was set to be freely estimated and the factor loadings on other factors were fixed to zero. Items with high loadings on more than one factor (differing <0.15) were set to load freely on both factors. To scale the estimated factors, on each factor the loading of one item was fixed to 1 (Brown, 2006). In the multi-factorial models, factor covariances were set to be freely estimated.

In the *1-factor model* all items loaded on one 'depression' factor (Trivedi et al., 2004). In the *2-factor model* (Bernstein, 2006) there was a 'depression' factor (all items) and a ‘somatic’ factor (items 25, 26 and 28). In the *3-factor model* (Rush et al., 1996), there was a ‘mood/cognition’ factor (items 5, 8, 10, 11-14 and 17-22), an ‘anxiety/ arousal’ factor (items 6, 7, 23-27 and 30) and a ‘sleep’ factor (items 1-4). Items 9, 15, 20 and 29 loaded on the ‘mood/cognition’ and ‘somatic’ factors and item 24 loaded on the ‘somatic’ and ‘sleep’ factors. In the *4-factor model* (Rush et al., 1986; 28 item version) there was a mood/cognition factor (items 5-8 and 15-22), an ‘anxious/hypochondria’ factor (items 24-28), an ‘endogenous’ factor (items 2, 3, 9, 11 and 13) and an ‘atypical’ factor (items 4, 12 and 14). Item 23 loaded on the first two factors, item 1 loaded on the second and third factor.

Because the IDS-SR items were categorical and had a non-normal distribution, estimation of model fit with maximum likelihood (ML) would likely lead to an underestimation of model-fit (Byrne, 2006). Therefore, we used an adapted approach for categorical data (Bentler, 2006). First, a matrix of polychoric correlations between the items was generated. Second, ML was used to estimate model fit-statistics. Third, the ML-based statistics were corrected using an appropriate weight-matrix to obtain *robust* fit-statistics (Satorra and Bentler, 1988), which have been shown to perform well for categorical and non-normal data (Byrne, 2006). Model-fit was evaluated with fit-indices, in stead of the traditional χ^2-test, which is oversensitive to minor deviations from perfect fit in large samples and with complex models (Brown, 2006). The following fit-indices were used: the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA) and the Akaike Information Criterion (AIC). A CFI of at least 0.95 indicates good fit and a RMSEA smaller than 0.06 indicates good fit (Hu and Bentler, 1999). The AIC can be used to compare different models, balancing statistical goodness-of-fit and the number of model parameters. The model with the lowest AIC can be regarded as potentially most useful (Bentler, 2006). In addition to the CFA in group 1, CFA was performed in groups 2, 3 and 4 with the best fitting model from group 1 to investigate the consistency of model-fit across different groups. The EQS statistical package (Multivariate Software Inc., Encino, California, USA) was used to conduct the analyses.

Rasch Analyses

To test if the identified factors were unidimensional measurement scales, fit to the Rasch model was investigated. The Rasch model assumes that the probability of a person’s response on an item is described by a *logistic* function of the distance between the location of the person and the location of the item on the underlying linear *severity dimension*. Thus, if a person is located higher on the dimension than the item, the probability of responding with the highest response option is high. Conversely, if the person is located lower than the item, the probability of responding with the lowest response option is high. If all items fit adequately to the Rasch model, this indicates that all items are lined up along one underlying dimension in order of increasing severity. In addition, fit to the Rasch model indicates that the ordinal responses on the items can be added up to a linear interval-scale that is a *sufficient statistic* for the underlying severity dimension.

The unrestricted partial credit model was used for fit-estimation. To estimate the fit to the Rasch model, the unweighted mean square standardized residual (outfit) was calculated for each item (formulas from: Wright & Masters, 1982, p100). Outfit was used in the current study because it is essentially a χ^2 statistic divided by its degrees of freedom (n), and thus less affected by the large sample size than traditional significance-tests of (mis)fit. An outfit close to 1 (within the range of 0.7 to 1.3) was considered to indicate adequate fit to the Rasch model (Wright & Stone, 1979). Persons with extreme scores (with a total score of 0 or with fit-residuals>|2.5|) were excluded from model-fit calculations because they do not behave in line with model expectations.

For each factor, the same analytic procedure was followed to assess and improve fit to the Rasch model. First, the fit of the items to the Rasch model was assessed. Second, the thresholds between adjacent response categories were inspected for ordering problems. If the response scale was disordered, for instance, if a category was never most probable to be endorsed (redundant), adjacent categories were collapsed. Third, the fit of the items was assessed again to see if fit had improved. Items with inadequate fit were removed to arrive at a unidimensional subscale with optimal fit to the Rasch model. The locations were inspected to see how the items were distributed along the underlying dimension. Fourth, all items in the final subscale were tested for differential item functioning (DIF) across person characteristics (Age group (young: 18-42 versus old: 43-65), gender and lifetime depression). An ANOVA was used to compare scores between levels of each person characteristic (uniform DIF) and across different classes of severity (non-uniform DIF). To investigate the actual extent and implications of DIF, the location and outfit were assessed for the different levels of the person characteristic (e.g. men vs. women). Because the ANOVA was likely to pick up less relevant DIF due to the large sample size (statistical power), a large difference in location (we chose >0.5 logits as cut-off) and model fit between two characteristic subgroups (e.g. men and women) was taken to indicate that the generalizability of item functioning is potentially problematic. Fifth, unidimensionality was additionally checked with a PCA of the residuals. Two subsets of items with respectively positive and negative loadings on the first component of this PCA were selected and person estimates were calculated for these subsets of items. If these estimates differed significantly from the full-scale estimates (as indicated by t>|1.96| in more than 5% of individuals), this indicated that the responses on the subsets were interdependent (when controlled for the primary underlying dimension) and that there was still some multidimensionality present in the measure (Smith, 2004). Sixth, to evaluate clinical usefulness of each subscale, the person-separation index was calculated and the number of severity strata that could be discriminated was derived from the separation-ratio (G). Additionally, to evaluate the (multi)dimensionality of the complete IDS-SR, its fit to the Rasch model was also investigated. Calculations were done with RUMM2020 (RUMM Laboratory, Perth, WA, Australia).

Missing Data

In group 1, 139 of 1096 subjects (12.7%) were excluded because they had one or more missing responses on the IDS-SR, resulting in a group of 957 subjects. In groups 2, 3 and 4, respectively 50 (10%), 48 (7.2%) and 58 (9.2%) subjects were excluded because of missing responses. The subjects with missing values were found to be younger and have less years of education than the subject with complete data (data not shown). However, it was decided not to impute missing items because to our knowledge there is no widely supported method to impute non-normal, categorical data without introducing new and unknown sources of bias.

Table 3.1: Demographic and diagnostic information of the studied subgroups in the NESDA data (n=2600)

Sample Diagnostic Status	Group 1 current MDD (<6 months)	Group 2 Remitted MDD (>6 months)	Group 3 Only anxiety (Lifetime)	Group 4 Healthy Controls
N	957	450	570	623
Women (%)	639 (66.8%)	319 (70.9%)	377 (66.1%)	382 (61.3%)
Men (%)	318 (33.2%)	131 (29.1%)	193 (33.9%)	241 (38.7%)
Mean age (SD)	40.5 (12.1)	43.8 (12.6)	41.7 (13.0)	40.8 (14.8)
Age range	18–64	18-65	18-65	18-65
Mean Yrs of Ed (SD)	11.7 (3.2)	12.5 (3.2)	12.3 (3.2)	12.9 (3.2)
DSM-IV diagnoses[a]				
Only current MDD[b]	331 (34.6%)	0	0	0
Only current anxiety	0	0	490 (86.0%)	0
Current comorbidity	626 (65.4%)	0	0	0
Only remitted MDD	0	301 (66.9%)	0	0
Only remitted anxiety	0	0	80 (14.0%)	0
Remitted comorbidity	0	149 (33.1%)	0	0

MDD = Major Depressive Disorder, Mean Yrs of Ed = mean years of education

[a] DSM-IV diagnoses assessed with the CIDI

[b] Also includes cases of current MDD + dythymia

3.3 Results

Diagnoses and demographic variables

The demographic and diagnostic information for the four non-overlapping study groups are shown in Table 3.1. The distribution of gender as well as mean age and mean years of education were largely comparable across the 4 groups. In group 1 (Current MDD), the majority (65.4%) of current MDD patients had a comorbid anxiety disorder. In group 2 (Remitted MDD), the majority (33.1%) of subjects had a comorbid remitted anxiety disorder. In group 3 (Lifetime anxiety disorders, no MDD), the majority (86.0%) of subjects had a current anxiety diagnosis. In group 4, by definition nobody had a psychiatric diagnosis.

Principal Component Analysis

The results of the initial PCA are shown in Table 3.2. Parallel analysis indicated that 3 components should be extracted. After rotation, items that covered symptoms of (depressed) mood, affect and cognitions loaded on the first component ('mood/cognition'), items that covered anxiety and somatic arousal and somatic complaints loaded on the second component ('anxiety/arousal') and items that covered sleep symptoms loaded on the third component ('sleep'). The extracted components were largely similar to the 3 components reported by Rush et al. (1996): only 7 of the 28 items had a completely different primary loading in the present study (items 6, 7, 15, 11/12, 13/14, 23 and 29). These results indicate that a 3-factor structure is likely to underlie the IDS-SR across a wide variety of subjects.

Confirmatory Factor Analysis

To test and compare the fit of the 4 models of the IDS-SR (1, 2, 3 and 4-facor), CFA was conducted in group 1 (Current MDD). The results are shown in Table 3.3. For the 3-factor model, the CFI was highest (CFI=0.95) and the RMSEA was lowest (RMSEA=0.056), which both indicated better fit than the 1, 2 and 4-factor models (all: CFI≤0.93; RMSEA≥0.065). In addition, the AIC was lowest for the 3-factor model (AIC=684.70), which indicates that this model is potentially most useful, taking into account both model-fit and the number of model-parameters. These results indicate that the 3-factor model proposed by Rush et al. (1996) best represents the underlying structure of the IDS-SR.

Confirmatory Factor Analysis in different groups

To investigate the consistency of model-fit of the 3-factor model, CFA was conducted in groups 2, 3 and 4. Results are shown in Table 3.4. The indices-of-fit indicated good model-fit in all groups (CFI≥0.95; RMSEA≤0.049), which indicated that the fit of the 3-factor model is consistent across subjects with remitted MDD, subjects with a lifetime anxiety disorder and healthy controls.

Table 3.2 Results of a Principal Components Analysis in the complete dataset (n=2600): factor loadings and eigenvalues

IDS-SR items	Component			Factor in Rush et al. (1996)[1]
	1. Mood/ Cognition	2. Anxiety/ Somatic	3. Sleep	
21. Pleasure or enjoyment (not sex)	**0.92**	-0.17	0.04	*1*
5. Feeling sad	**0.89**	-0.09	0.02	*1*
8. Reactivity of mood	**0.86**	-0.26	0.06	*1*
19. Interest in people/activities	**0.84**	-0.10	0.01	*1*
17. Future pessimism	**0.78**	-0.06	0.03	*1*
10. Quality of mood	**0.72**	-0.01	-0.03	*1*
16. Self criticism and blame	**0.66**	0.02	-0.10	*1*
15. Concentration/decision making	**0.64**	0.14	0.01	*2*
18. Suicidal thoughts	**0.60**	-0.07	0.01	*1*
6. Feeling irritable	**0.59**	0.14	0.01	*2*
23. Psychomotor retardation	**0.57**	0.10	0.04	*2*
20. Energy/fatiguability	**0.57**	0.27	-0.05	*1*
22. Interest in Sex	**0.53**	0.05	0.08	*1*
29. Interpersonal sensitivity	**0.51**	0.20	-0.16	*2*
7. Feeling anxious or tense	**0.50**	0.32	0.03	*2*
30. Leaden paralysis/physical energy	**0.46**	**0.39**	-0.04	*2*
28. Constipation/diarrhoea	-0.22	**0.79**	-0.03	*2*
25. Aches and pains	-0.06	**0.69**	0.14	*2*
26. Sympathetic arousal	0.08	**0.58**	0.13	*2*
27. Panic/phobic symptoms	0.08	**0.56**	-0.04	*2*
13/14. Weight disturbance	-0.13	**0.53**	0.01	*1*
11/12. Appetite disturbance	0.23	**0.43**	-0.09	*1*
24. Psychomotor agitation	**0.25**	**0.29**	0.10	*2*
3. Early morning awakening	0.12	0.00	**0.69**	*3*
2. Middle insomnia	-0.06	0.19	**0.68**	*3*
4. Sleeping too much	0.15	0.31	**-0.56**	*3*
1. Initial insomnia	0.16	0.19	**0.42**	*3*
9. Diurnal variation of mood	0.16	0.16	-0.04	-
Eigenvalue (in real data)	9.97	1.47	1.26	
Eigenvalue (randomly generated)	1.22	1.18	1.16	

Table 3.2 (continued). Legend: IDS-SR = Inventory of Depressive Symptomatology Self Report; communalities after extraction ranged from 0.21 to 0.70; the components were rotated with PROMAX; the primary loading for each item is printed in bold font

[1)] Components on which each item had its highest loading in PCA results by Rush et al. (1996): 1 = 'mood/cognition', 2= 'anxiety/arousal', 3 = 'sleep'; an underlined number indicates that the item loads on a different component in the present study

Table 3.3: Confirmatory factor analysis of four factor-models of the IDS-SR in group 1 (Current MDD; n = 957)

Input model	Source	Df	Satorra-Bentler χ^2	CFI	RMSEA (90% CI)	AIC
1-factor	Trivedi et al. (2004)	350	2043.23	0.92	0.071 (0.068 – 0.074)	1343.23
2-factor	Bernstein (2006)	347	1845.39	0.93	0.067 (0.064 -0.070)	1151.39
3-factor	Rush et al. (1996)	343	**1370.70**	**0.95**	**0.056 (0.053 – 0.059)**	**684.70**
4-factor[a]	Rush et al. (1986)	291	1466.78	0.93	0.065 (0.061 - 0.068)	880.78

IDS-SR = Inventory of Depressive Symptomatology Self Report; Df = Degrees of freedom; CFI = Comparative fit index; Standardized root mean-square residual; RMSEA = Root mean square error of approximation; 90% CI (RMSEA) = 90% confidence interval of the RMSEA; AIC = Akaike Information criterion

CFA based on polychoric correlation matrix with robust Satorra-Bentler correction

[a] The 4-factor input model included 26 of the 28 IDS-SR items: of the 4 appetite/weight items, only item 11 (decreased appetite) and item 14 (increased weight) were included because the polychoric correlations between item 11 and item 12 (increased appetite) and between item 13 and item 14 (decreased weight) both approached -1.0, resulting in a non-positive definite correlation matrix, which can not be used to estimate model-fit. By including only one item of each pair in the model, this problem was solved and the proposed distinction between an atypical and endogenous factor was still expressed in the model

Rasch Analyses

The full IDS-SR

Of the 28 items in the IDS-SR, 10 items showed poor fit to the Rasch model, which is further evidence for the multidimensionality of the IDS-SR.

The IDS-SR 'Mood/Cognition' subscale

All 15 items that were set to load on the 'Mood/Cognition' factor in the CFA were investigated for fit to the Rasch model. Items 9, 13/14, 19 and 21 did have an outfit statistic outside the acceptable range. Inspection of the threshold-ordering revealed that seven items (9, 11/12, 14-17 and 29) had disordered thresholds, which resulted from category 1 or 2 to be redundant in each of these items. Therefore, it was decided to collapse category 1 and 2, resulting in a 3-point response scale (0=0, 1=1, 2=1 and 3=2). For ease of use all items were rescored accordingly. The thresholds, locations and outfit after rescoring are shown in Table 3.5. All thresholds were now ordered correctly and item-fit had mostly increased after rescoring. However, the same four items still fit poorly and were therefore removed. This resulted in a final 11-item subscale with item-locations ordered as follows (in ascending order): 29, 16, 17, 10, 15, 20, 5, 22, 11/12, 8 and 18.
Thus, interpersonal sensitivity (item 29), problematic self view (item 16) and pessimism about the future (item 17) were at the low end of the 'mood/cognition' dimension and decreased reactivity of mood (item 8) and suicidal thoughts (item 18) were at the severe end.

Several items displayed DIF (significant DIF and difference between locations>0.5). Item 5 displayed DIF between depressed (location=-0.24) and non-depressed subjects (location=0.51) but had adequate outfit in both subgroups (0.7<outfit<1.3). Item 8 displayed DIF between men (location=0.66) and women (location=1.37) but showed adequate outfit in both groups. Item 29 displayed DIF between men (location=-0.55) and women (location=-1.09) and between young (location=-1.23) and old subjects (location=-0.60) but had adequate outfit in all subgroups. Inspection of the location-ordering indicated that the items were generally ordered in the same way in the different person-factor groups, with items 29, 16 and 17 at the lowest and items 8 and 18 at the highest end of the severity dimension. Thus, although items 5, 8 and 29 show some DIF, the consistent adequate outfit and threshold ordering across subgroups indicates this does not severely affect the generalizability of measurement.

Comparing the person estimates between the PCA item-subsets and the final 11 item 'mood/cognition' subscale indicated no significant difference for any person (all t-values <|1.96|), which forms further evidence for unidimensionality.

The 11-item 'mood/cognition' subscale had a person-separation-index of 0.88, which indicates that the scale can be used to discriminate around 4 severity strata (G≈3; Wright & Masters, 1982). The removal of 4 items from the initial 15-item subscale only led to a 0.01 reduction of the person-separation-index, indicating that these items did not contribute substantially to the discriminative ability of the subscale.

Table 3.4: Confirmatory factor analysis of the 3-factor model of the IDS-SR in group 2 (remitted MDD), group 3 (Lifetime anxiety) and group 4 (Healthy controls)

Model	Group	Df	Satorra-Bentler χ^2	CFI	RMSEA (90% CI)
3-factor	2 (remitted MDD), n = 450	343	670.34	0.96	0.046 (0.041 – 0.051)
	3 (Anxiety), n = 570	343	821.17	0.95	0.049 (0.045 – 0.054)
	4 (healthy Controls), n = 623	343	752.14	0.96	0.044 (0.040 – 0.048)

IDS-SR = Inventory of Depressive Symptomatology Self Report ; MDD = Major Depressive Disorder; Df = Degrees of freedom; CFI = Comparative fit index; Standardized root mean-square residual; RMSEA = Root mean square error of approximation; 90% CI (RMSEA) = 90% confidence interval of the RMSEA
CFA based on polychoric correlation matrix with robust Satorra-Bentler correction.

The IDS-SR 'Anxiety/Arousal' subscale

All nine items that were set to load on the 'anxiety/arousal' factor in the CFA were investigated for fit to the Rasch model. Only item 7 fit poorly (outfit=0.63). Inspection of the threshold-ordering again revealed redundancy of category 1 or 2 in some items (items 23, 24 and 30). Therefore, the items in the 'anxiety/arousal' subscale were also rescored to a 3-point scale (0=0, 1=1, 2=1 and 3=2). The thresholds, locations and outfit statistics are shown in table 3.5. All thresholds were now ordered correctly and item fit had mostly increased after rescoring. However, item 7 still fit poorly and was removed. This resulted in an 8-item subscale with item-locations ordered as follows: 30, 25, 6, 27, 28, 24, 26 and 23. Thus, symptoms such as leaden paralysis (item 30), aches and pains (item 25) and feeling irritable (item 6) were at the lower end of the severity dimension and psychomotor agitation (item 24), sympathetic symptoms (e.g. arrhythmic or pounding heartbeat, blurred vision, sweating; item 26) and psychomotor retardation (item 23) were on the severe end.

Items 24 and 25 displayed DIF between men (location=-0.03 and -0.23) and women (location=0.58 and -0.77) but both had adequate fit in both subgroups (0.7<outfit<1.3). Item 6 displayed DIF between young (location=-0.53) and old (location=-0.05) but had adequate fit in both groups. Item 27 displayed DIF between depressed (location=0.34) and non-depressed (location=-0.20) and showed poorer fit in the depressed (outfit=0.69) than in the non-depressed subgroup (outfit=0.86). Inspection of the location-ordering indicated that the items were generally ordered in the same way in the different person-factor groups, with items 30, 25 and 6 at the lowest and items 24, 26 and 23 at the highest end of the dimension. The results indicate that items 24 and 25 showed some DIF that was not likely to severely affect the generalizability of measurement. The DIF for item 27 was more serious, although its outfit statistic of 0.69 in depressed subjects was only just below the cut-off of 0.70, the function of this item could be less consistent.

3.4 Discussion

The present study used CFA to identify the best fitting structural model of the IDS-SR. An initial PCA indicated that 3 factors were expected to underlie the IDS-SR and when the fit of four models from the literature was tested with CFA, the 3-factor model (Rush et al., 1996) was indeed found to fit best to the data. This model consisted of a 'mood/cognition', 'anxiety/arousal' and 'sleep' factor and was found to fit well across different groups of patients and healthy controls. To evaluate if they could be used as subscales, the factors were tested and fine-tuned using Rasch analyses: items were rescored to a more optimal 3-point scale (0, 1, 1, 2) and items that fit the model poorly were removed. This resulted in two unidimensional IDS-SR subscales: the 11-item 'mood/cognition' subscale and the 8-item 'anxiety/arousal' subscale. The adequate fit to the Rasch model indicated that the sum scores on these subscales can be regarded as *sufficient statistic* for their underlying symptom dimensions. An additional PCA of the residuals indicated that the scales were unidimensional. DIF analyses showed that some items functioned differently across groups, though measurement characteristics generally seemed consistent. Finally, the subscales were found to have adequate discriminative ability. Importantly, Rasch analyses with the total IDS-SR indicated that it is multidimensional, underlining the need for more homogeneous symptom measures.

The IDS-SR subscales could be helpful for both clinicians and researchers who seek less heterogeneous symptom-severity measures. Using the subscales could have added value because: (1) they function well as 'measurement scales' (as shown by the Rasch results), (2) they assess severity for more specific symptom-domains and (3) different patterns of subscale scores could indicate different treatment indications and/or disease prognosis. In addition, the finding of two symptom dimensions could indicate (partly) distinct underlying etiological mechanisms.The present results have some additional general implications. First, the traditionally used total sum score on the IDS-SR does not seem to be a unidimensional measure of depression-severity. This is a general problem that is also observed for other widely used instruments like the Hamilton Depression Rating Scale (HDRS; Gibbons et al, 1993) the Beck Depression Inventory (BDI; Gibbons et al., 1985) and the Centre for Epidemiological Studies Depression Scale (CES-D; Stansbury et al., 2006). Second, the finding that a 3-dimensional model fit the IDS-SR better than the 4-dimensional model (Rush et al., 1986), does contradict the traditional idea of distinct 'atypical' and 'melancholic/endogenous' symptom domains.

Table 3.5: The location and outfit statistic for each of the items in the IDS-SR 'Mood/Cognition' subscale and 'Anxiety/Arousal' subscale

	IDS-SR 'Mood/Cognition' subscale Threshold				IDS-SR 'Anxiety Arousal' subscale Threshold			
	1	2	Location	Outfit	1	2	Location	Outfit
5. Feeling sad	-2.38	2.09	-0.15	**0.72**	-	-	-	-
6. Feeling irritable	-	-	-	-	-2.56	1.99	-0.28	**0.84**
7. Feeling anxious or tense	-	-	-	-	-2.61	2.13	-0.24	0.67
8. Reactivity of Mood	-0.56	2.50	-0.97	**1.01**	-	-	-	-
9. Diurnal variation of mood	-0.50	0.60	0.05	2.26	-	-	-	-
10. Quality of mood	-1.36	0.56	-0.40	**0.79**	-	-	-	-
11/12. Appetite disturbance	-1.15	1.01	-0.07	**1.04**	-	-	-	-
13/14. Weight disturbance	-1.42	0.89	-0.26	2.31	-	-	-	-
15. Concentration/decision making	-2.37	1.79	-0.29	**0.82**	-	-	-	-
16. Self criticism and blame	-1.07	-0.46	-0.76	**0.96**	-	-	-	-
17. Future pessimism	-2.90	2.06	-0.42	**0.82**	-	-	-	-
18. Suicidal thoughts	-3.42	2.92	0.29	**1.01**	-	-	-	-
19. Interest in people/ activities	-1.04	1.59	0.27	0.62	-	-	-	-
20. Energy/fatiguability	-2.26	1.96	-0.15	**0.82**	-	-	-	-
21. Pleasure or enjoyment	-1.17	3.08	0.96	0.60	-	-	-	-
22. Interest in sex	-1.37	1.19	-0.09	**1.26**	-	-		-
23. Psychomotor retardation	-	-	-	-	-0.34	2.36	1.01	**0.72**
24. Psychomotor agitation	-	-	-	-	-1.36	2.03	0.33	**1.01**
25. Aches and pains	-	-	-	-	-3.07	1.91	-0.58	**0.96**
26. Sympathetic arousal	-	-	-	-	-2.52	3.81	-0.65	**0.85**
27. Panic/phobic symptoms	-	-	-	-	-1.66	1.92	0.13	**0.96**
28. Constipation/diarrhoea	-	-	-	-	-1.45	1.78	0.16	**1.19**
29. Interpersonal Sensitivity	-2.45	0.55	-0.95	**1.01**	-	-	-	-
30. Leaden Paralysis	-	-	-	-	-2.85	0.49	-1.18	**0.76**

IDS-SR = Inventory of Depressive Symptomatology Self Report. Items 1 to 4 are omitted from the table. Threshold 1= location of threshold between response option 0 and 1; threshold 2=location of threshold between response option 1 and 2; Location = value between threshold 1 and 2. Items with an outfit in bold font were retained in the subscale

Evidence of heterogeneity has been found for many widely used depression scales (e.g. the HDRS, BDI and CES-D), and many shorter, unidimensional versions and/or subscales have been proposed. For the HDRS, subsets of items have been proposed to measure only ‘core symptoms’ of depression (Bech et al, 1981; Maier and Philipp, 1985; Santor et al., 2008). These ‘subscales’ were indeed shown to function as unidimensional measures with IRT analyses (e.g. Bech et al., 1981; Gibbons et al., 1993; Santor et al, 2008). Recently, the same was found for the self rated HDRS version (Bech et al., 2009). For the BDI, several revisions and subscales have been proposed. For instance, Gibbons et al (1985) found the BDI to be more unidimensional without vegetative symptoms. Bouman & Kok (1987) further subdivided the BDI into three unidimensional subscales using Rasch analyses: ‘mood/inhibition’, ‘guilt/failure’ and ‘vegetative’. This subdivision is in line with the most commonly found factor structure for the BDI (review: Beck et al., 1988; Shafer, 2006). Also for the CES-D shorter (IRT derived) unidimensional subscales have been developed (e.g. Cole et al., 2004; Stansbury et al., 2006). The current IDS-SR ‘mood/cognition’ subscale can be regarded as conceptually similar to the abovementioned attempts to create a more homogenous depression measure. However, coverage seems to vary somewhat across the different subscales, most likely because each original instrument has a slightly different focus and item-pool to select from. The IDS-SR ‘anxiety/arousal’ subscale mostly resembles the item-sets that have been found to load on one ‘anxiety-agitation’ factor or two distinct ‘somatic anxiety’ and ‘psychic anxiety’ factors in the HDRS (Bagby et al., 2004; Shafer et al., 2006). However, these factors have received less attention, which is not surprising given that the instrument is mainly used as a depression measure in antidepressant trials. Both the BDI and CES-D mainly focus on depressed mood and/or cognitions and cover only a few somatic/vegetative symptoms, which is not enough to construct a reliable subscale (Gibbons et al., 1985). The present finding of a ’sleep’ factor is in line with other studies that found sleeping problems to load on a separate ‘insomnia’ factor (HDRS) or on a ‘neurovegetative’ factor (BDI; CES-D) (Shafer, 2006).

The present study has several strong characteristics. First, CFA was conducted in large and representative samples of healthy subjects and psychiatric patients, which makes the results generalizable to a broad range of settings and patients. Second, CFA was used to test the fit of multiple hypothesized models, which allowed both for assessment of model fit and selection of a best-fitting model from among several plausible models. Third, a CFA approach was used that minimized bias due to the categorical and non-normal nature of the IDS-SR data, increasing the validity of the results. Fourth, the Rasch analyses resulted in subscales that can really be regarded as unidimensional measurement scales; something that would not have been achieved with classical psychometric analyses. The results should be interpreted in the light of some limitations. First, only the fit of four models from the literature was tested, which does not rule out that an unknown model with a different structure (e.g. more factors) might fit better. However, the initial PCA in the present study did not suggest that this was the case. Second, the current analyses were conducted in a group of subjects with relatively mild psychopathology. Future research should point out whether our findings can also be generalized to more severely ill and/or institutionalized patients. Finally, although DIF did not seem to be very problematic for general use, researchers that are specifically interested in score differences between groups (e.g. men and women) should be aware of DIF that could reduce subscale-score comparability. They could leave DIF items out of the subscale calculations.

In conclusion, the IDS-SR has three underlying factors, of which two can be adapted for use as specific subscales in both clinical practice and scientific research.