First publ. in: Applications of Latent Trait and Latent Class Models in the Social Sciences / Jürgen Rost and Rolf Langeheine (eds.). Münster: Waxmann, 1997, pp. 314-322

Some Latent Trait and Latent Class Analyses of the Beck-Depression-Inventory (BDI)

Ferdinand Keller and Wilhelm Kempf

Zentrum für Psychiatrie Weissenau (University of Ulm)
University of Konstanz

1. Introduction

For assessing the intensity of depression in both psychiatrically diagnosed and normal populations, the Beck Depression Inventory (BDI) is one of the most widely used instruments. The BDI was introduced in 1961 and revised in 1978 (Beck, Steer & Garbin, 1988). It is a self-report questionnaire and consists of 21 items derived by clinical observations about symptoms displayed frequently by depressed patients and infrequently by nondepressed psychiatric patients. The items were chosen to assess the intensity of depression and were not selected to reflect a particular theory of depression (Beck, Steer & Garbin, 1988). Each item contains four statements reflecting increasing levels of intensity of a depressive symptom. The items are scored from 0 to 3 and are summed for an overall depression score.

Evaluation of the psychometric properties of the BDI and tests for the presence of one or more underlying dimensions were mostly performed within the framework of classical test theory, correlational analysis and factor analysis. The psychometric properties are considered to be good (Beck, Steer & Garrison, 1988), and Hautzinger (1991) confirmed the main results for a German translation of the BDI. A recent overview by Richter, Werner, and Bastine (1994) discusses further research on these topics. Factor analyses of the BDI have resulted in several different solutions (Richter, 1991). However, a general factor has been found in many studies (and for the German translation (Hautzinger, 1991)), and, therefore, the BDI is assumed to represent one underlying general syndrome of depression that can be decomposed into three highly intercorrelated factors (Beck, Steer & Garbin, 1988).

Studies using probabilistic models to test unidimensionality are scarce. Clark et al. (1983, 1985) applied the Birnbaum model, and Bouman & Kok (1987) performed a dichotomous Rasch analysis of the BDI. They revealed that the BDI does not, in terms of the Rasch-Model, constitute a unidimensional scale for the measurement of depression in clinical samples. Nonetheless, the authors proposed three subscales of the BDI that were found to be homogeneous (»Guilt & Failure«, »Mood & Inhibition« and »Somatic« (cf. Table 1)) and intercorrelated modestly.

In this paper, we give a short summary of our previous research concerning the replication of homogeneity in a German sample and the extension to polytomous Rasch analysis and latent class analysis (LCA) of the Bouman-Kok subscales. Then, we show results of the application of LCA to analyze second order relationships of class-membership on these three

subscales. The contingency of the syndromes (identified by second order LCA) with gender and time of assessment is analyzed as well. Finally, we propose several other subscales derived from the BDI (the main symptoms of depression according to DSM-III-R and three subscales composed with regard to clinical importance and similarity) and discuss their homogeneity.

2. Methods

The sample consisted of BDI data for depressive inpatients treated at two depression wards in psychiatric state hospitals. These 400 patients supplied results for 714 BDIs. The percentage of BDIs filled in by woman is 70.6%, and the distribution for the time of assessment is 42.3% at admission, 39.1% at discharge, and 18.6% at follow-up (one year after discharge).

Item (#)	*Symptom*	*Boumans& Kok*[a]	*DSM-III-R (SKID)*[b]	*clinical scales*	*symptom present at*[c]
A (1)	Sadness	M & I	xx	affective	2
B (2)	Pessimism	M & I		cognitive	2
C (3)	Sense of failure	G & F		cognitive	2
D (4)	Dissatisfaction	-		-	
E (5)	Guilt	G & F	x	cognitive	1
F (6)	Sense of punishment	-		-	
G (7)	Self-hate	G & F		-	
H (8)	Self-blame	G & F		cognitive	2
I (9)	Suicidal ideation	M & I	x	cognitive	1
J (10)	Crying spells	SOM		affective	2
K (11)	Irritability	SOM		affective	2
L (12)	Loss of social interest	-	xx	-	2
M (13)	Indecision	M & I	x	cognitive	2
N (14)	Poor body image	-		-	
O (15)	Work inhibition	M & I		-	
P (16)	Sleep disturbance	SOM	x	vegetative	2
Q (17)	Fatigue	SOM	x	vegetative	2
R (18)	Loss of appetite	SOM		vegetative	2
S (19)	Weight loss	SOM	x	vegetative	1
T (20)	Hypochondriasis	SOM		-	
U (21)	Loss of libido	SOM		vegetative	2

[a] G & F: guilt & failure; M & I: mood & inhibition; SOM: somatic

[b] xx = core symptom; x = additional symptom

[c] number indicates first response category treated as „symptom present" (= 1), e.g., the number 2 indicates a dichotomization into x = 0,1 vs. x = 2,3. For the Bouman & Kok subscales (and our replication thereof) the dichotomization is always x = 0 vs. x > 0.

Table 1: Symptom description of the BDI-Items, their membership to different subscales and thresholds for dichotomization.

The items of the BDI and the composition of the various subscales are listed in Table 1. All items are dichotomized for the following analyses: those of the Bouman & Kok-subscales as proposed in their original paper and those for the other subscales are divided at the diagnostically relevant threshold according to DSM-III-R or clinical experience (cf. Table 1). For the clinical subscales, the tests for unidimensionality with the mixed Rasch model

approach (Rost, 1990) are followed by evaluation of models with weaker assumptions (LCA) to explore possible violations of homogeneity (see section 2.1 in chapter 1). Data were analyzed by means of the programs MIRA and LACORD (provided by J. Rost) and WINMIRA (v. Davier, 1994).

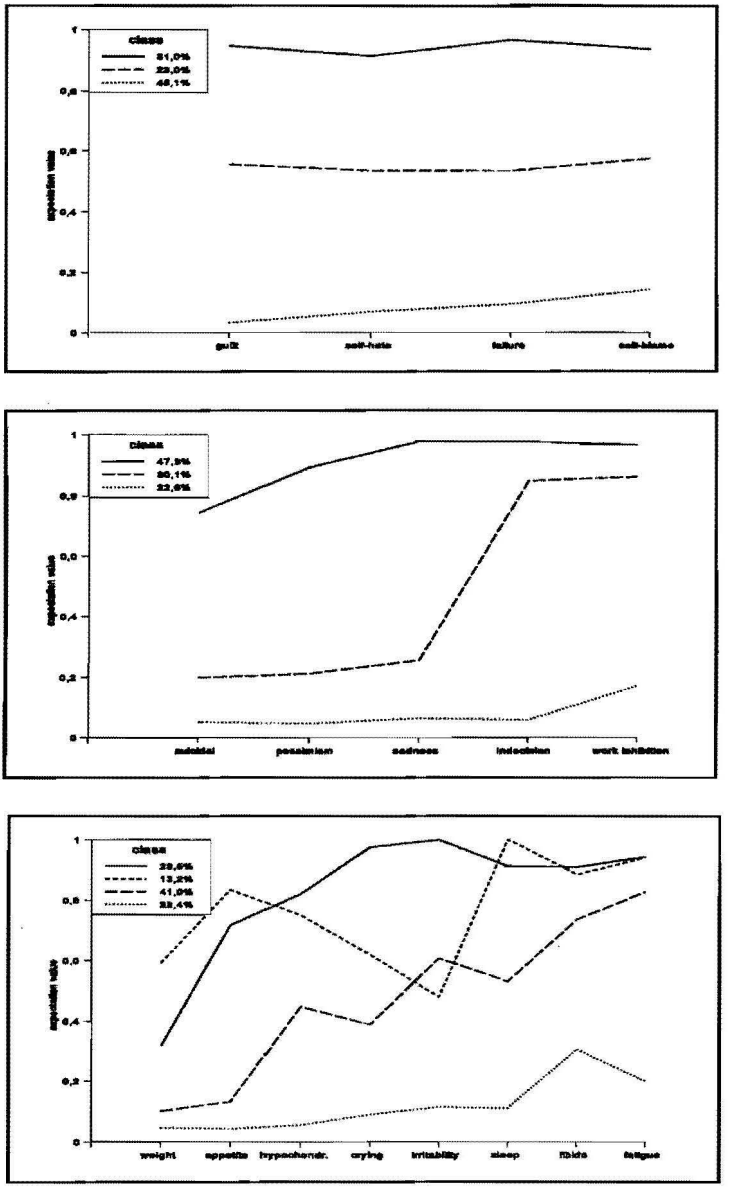


Figure 1: Expectation values for the Bouman & Kok-subscales (symptoms dichotomized (cf. Table 1), and arranged according to increasing item difficulties)

3. Summary of our previous research

The analysis of the Bouman & Kok-subscales in the German version of the BDI revealed that they did not meet the strong assumptions of the Rasch model regardless of whether the item scores were dichotomized or the original polytomous scores were used (Kempf & Keller, 1992). Nonetheless, LCA of the subscales identified latent classes of subjects that did not show strong qualitative differences of response patterns, but could essentially be described by a quantitative graduation of strength and propensity of symptoms (cf. Figure 1 for the patterns with dichotomized items). Furthermore, the identification of the latent classes was practically independent of gender and time of assessment (Keller & Kempf, 1993). Therefore, it is concluded that the subscales have satisfactory psychometric properties.

4. Results for second order LCA

The latent classes obtained by LCA on each of these three subscales can be inserted as variables into a second order LCA, with the most likely class-membership as the categories coded on the respective variable. The second order LCA revealed that an over-all homogeneity of the BDI cannot be confirmed (Table 2). The 4-classes solution shows the best fit according to the AIC, although the BIC would favor a 3-classes solution.

number of classes	*ln(L)*	*number of parameters*	*AIC*	*BIC*
1	-2392.26	7	4798.53	4830.52
2	-2065.05	15	4160.09	4228.66
3	-2003.49	23	4052.09	4158.12
4	-1989.81	31	4041.61	4183.31

Table 2: Model tests for second order LCA.

Inspection of the expectation values shows qualitatively defined syndromes of class-membership on the different scales (Figure 2). Especially, class #3 (size 9.5%) is characterized by a high somatic component, in connection with the absence of guilt/failure.

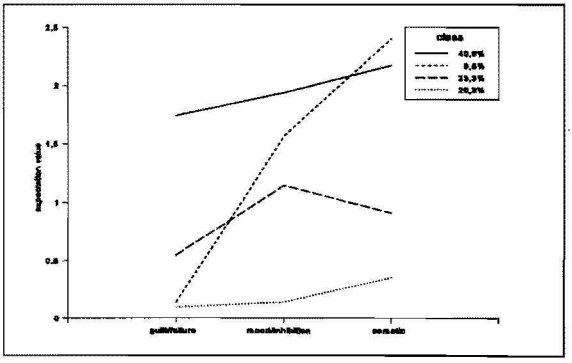


Figure 2: Expectation values for the second order LCA

In order to explore the contingencies between latent classes and external variables, e.g., gender, the conditional class sizes within the respective subsamples can be obtained as the mean membership probabilites within the male and female subgroup. The same method was applied to the evaluation of contingencies with the time of assessment. The contingency of the second order latent classes with gender is very low (Table 3), i.e., males and females do not significantly differ with respect to class-membership.

class	*male (0.294)*	*female (0.706)*
1	0.2455	0.2707
2	0.2480	0.2268
3	0.0799	0.1007
4	0.4265	0.4018

Table 3: Conditional class sizes for male and female subgroup (chi^2=1.54; df=3; n.s.).

With regard to the time of assessment (Table 4) the class sizes are strongly different (as expected), e.g., the class #4 (high symptomatic) is 60.8% at admission and only 30.8% at discharge. The qualitatively different class #3 is particularly found at admission (16%), and it diminishes later on (6% at discharge, 2.5% at follow-up). Interestingly, the low symptomatic classes (#1 and #2) are even more pronounced at follow-up than at discharge, and the opposite is true for the symptomatic classes (#3 and #4).

class	*admission (0.423)*	*discharge (0.391)*	*follow-up (0.186)*
1	0.0789	0.3623	0.4740
2	0.1564	0.2694	0.3307
3	0.1566	0.0605	0.0252
4	0.6080	0.3077	0.1700

Table 4: Conditional class sizes for the time of assessment (chi^2=163.74; df=6; p<.001).

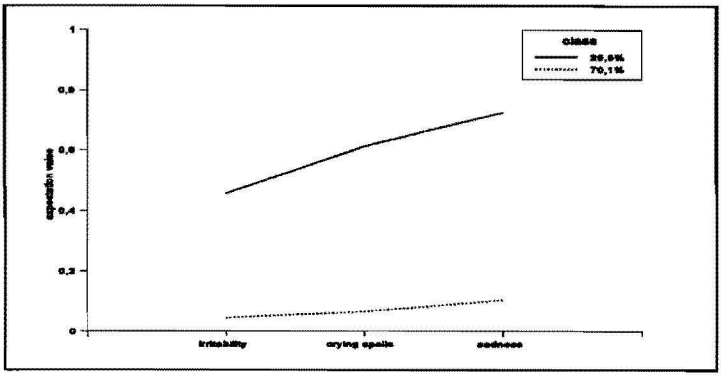
5. Clinical Subscales

As an alternative to the Bouman & Kok-subscales a number of subscales representing different clinical aspects of depression was composed (cf. Table 1). Additionally, the items of these subscales were dichotomized according to diagnostically relevant thresholds which are preferable to a simple $x = 0$ vs. $x > 0$ criterion.

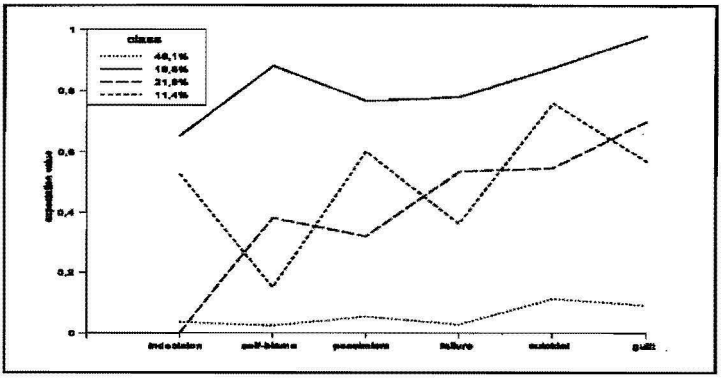
5.1 Cognitive subscale

The Rasch analysis of the cognitive subscale reveals no clear picture: the AIC for the one-class and the two-classes-solution is almost the same, while the BIC favors the one-class solution (Table 5). Further analyses with less restrictive assumptions suggested a 4-classes LCA-solution (again, the BIC differs from the AIC). According to the profile lines obtained by the LCA (Figure 3), one class (size 11.4%) is qualitatively different. In comparison with the other „middle class", it is characterized by higher indecisiveness, pessimism, and suicidal wishes, and lower self-blame, sense of failure, and guilt. A very speculative interpretation might deal with a time distinction: the first class is more oriented to a (negative) future, the latter is looking back.

affective



cognitive



vegetative

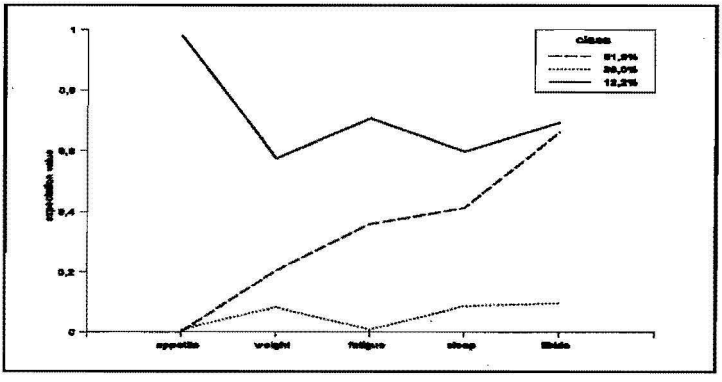


Figure 3: Expectation values for the clinical subscales (symptoms dichotomized (cf. Table 1), and arranged according to increasing item difficulties)

Across all models fitted (Rasch, LCM), both AIC and BIC favor the simple Rasch model; perhaps this is the most appropriate model for the cognitive subscale. Also, the LR chi-square test for the simple Rasch model versus the 4-classes LCM is 25.50 (df = 16) which is non-significant suggesting Rasch homogeneity.

model	*number of classes*	*ln(L)*	*number of parameters*	*AIC*	*BIC*
mixed	1	-2178.68	11	4379.36	4429.64
Rasch	2	-2168.69	21	4379.38	4475.37
	3	-2161.30	31	4384.60	4526.30
latent	1	-2639.42	6	5290.83	5318.27
class	2	-2199.68	13	4425.36	4484.78
	3	-2173.33	20	4386.67	4478.09
	4	-2165.93	27	4385.86	4509.27
hybrid	2	-2172.09	18	4380.18	4462.45

Table 5: Comparison of different models for the clinical subscale „cognitive" (6 items).

An alternative hypothesis would be that the cognitive subscale is homogeneous within the majority of the clinical population, with a minority of "unscalables", as described by the hybrid Rasch-LC model (cf. eq. (50) in chapter 1 and chapter 28 by v. Davier & Rost). Applying this model, we found a large Rasch-homogenous class (87.8%) and a small additional latent class (12.2%). Although the size of this latent class is close to the size of the qualitatively different latent class in the LCA, the two models do not identify the same subclass. The profile line of the latent class obtained by the hybrid model is similar to the class with size 21.9% in the LCA-solution.

5.2 Vegetative subscale

The vegetative subscale cannot be considered to be Rasch-homogenous (Table 6), because the AIC suggests a 2-classes solution (the BIC differs again). The LCA reveals consistently a 3-classes solution with better AIC and BIC than the MIRA-solutions. The comparison of the Rasch model and the 3-classes LCM by means of the LR test is highly significant (chi-square = 54.54, df = 8) suggesting also that the items are not Rasch homogenous.

model	*number of classes*	*ln(L)*	*number of parameters*	*AIC*	*BIC*
mixed	1	-1875.12	9	3768.25	3809.38
Rasch	2	-1850.66	17	3735.31	3813.03
	3	-1844.12	25	3738.25	3852.51
latent	1	-1984.30	5	3978.60	4001.45
class	2	-1868.98	11	3759.97	3810.24
	3	-1847.85	17	3729.71	3807.41
mixed Rasch					
(excluding	1	-1656.51	7	3327.02	3359.02
„appetite")	2	-1653.21	13	3332.43	3391.84

Table 6: Comparison of different models for the clinical subscale „vegetative" (5 items).

An inspection of the expectancy values (Figure 3) immediately shows that the discrepancy in the item „loss of appetite" is responsible for a qualitatively different class. The omission of this symptom reveals a Rasch-homogeneous subscale (Table 6, lower part).

5.3 Affective subscale

The affective subscale consists of only three items. Therefore, the estimation of mixed Rasch models is possible for one class only. The one-class solution reveals an AIC = 2098.04 and BIC = 2120.89. The LCA favors a 2-classes solution (Table 7). The profile lines for the two classes show no violation of order relations (Figure 3). Moreover, the AIC and BIC for the LCA exceed those for the Rasch model. This also suggests Rasch homogeneity.

subscale	*number of classes*	*ln(L)*	*number of parameters*	*AIC*	*BIC*
affective	1	-1136.75	3	2279.50	2293.21
	2	-1043.94	7	2101.88	2133.88
DSM-III-R	1	-3289.76	8	6595.52	6632.09
	2	-2826.22	17	5686.44	5764.15
	3	-2790.60	26	5633.20	5752.04
	4	-2778.23	35	5626.46	5786.44

Table 7: Model tests (LCA) for the clinical subscales „affective" (3 items) and „DSM-III-R" (8 items).

5.4 DSM-III-R subscale

This combination of items is not meant to be a homogeneous scale in assessing a major depressive episode (MDE) according to the DSM-III-R. The diagnosis of a MDE requires the presence of five symptoms out of nine and at least one core symptom. The BDI reflects six of the nine DSM-III-R criteria well, but two are „one-sided" (loss of appetite and sleep difficulty, excluding increased appetite and sleeping). One item (psychomotoric agitation or retardation) is missing because this is a clinically observable sign, not appropriate for self-reporting. Therefore, only the LCA is applied in order to look for patterns in the main symptoms of depression (of course, each subscale of the BDI should be homogeneous, too, if the BDI measures one latent dimension). The AIC suggests a 4-classes-solution (BIC is deviant, again; Table 7).

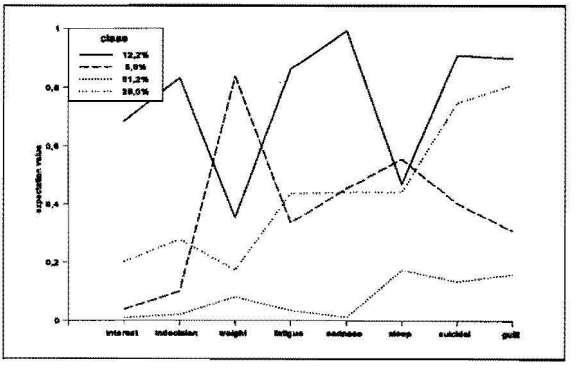


Figure 4: Expectation values for the subscale „DSM-III-R" (symptoms dichotomized (cf. Table 1), and arranged according to increasing item difficulties)

The crossing of the profile lines (Figure 4) indicates a violation of the homogeneity assumption. One class (size 8.6%) is qualitatively different, while the other ones are approximately compatible with regard to their order relations. Therefore, the assumption of an homogeneous depression syndrome is not tenable. It has to be repeated, however, that the DSM-III-R diagnosis is based on the clinician's rating, while the BDI is self-administered, and the item formulations are not exactly the same.

5.5 Preliminary results for polytomous analysis

For reasons of space, we cannot go into details for the analysis with polytomous items, but the main interpretations tend to be quite obvious. We concentrated on the cognitive subscale, because there are previous results with cognitive symptoms to be more predictive for the time to relapse („survival rate") than somatic and affective symptoms at the time of discharge (Keller et al., 1991).

Polytomous Rasch analysis of the cognitive subscale revealed that a two-classes solution is the most appropriate according to AIC (the BIC suggests Rasch homogeneity). Both classes have about the same size, but only one of them has ordered threshold parameters, i.e., the sequence of item categories does not reflect increasing levels of the latent dimension for one class in five items of the subscale. Threshold parameters were allowed to vary freely within each item (i.e., a partial credit model was estimated), and this model has proved to be necessary, because models with restrictions on the category parameters (e.g., the rating scale model which assumes the threshold differences to be constant for all items) showed markedly increased AICs und BICs.

6. Discussion

The main results can be summarized as follows. For the dichotomized version, the second order LCA (LCA of class-membership categories for the Bouman & Kok-subscales) and the subscale reflecting the DSM-III-R symptoms revealed that the BDI does not fulfill the strong requirements of Rasch homogeneity (in agreement with Bouman & Kok, 1987). Explorative LCA, however, confirmed the violation to be limited to a relatively small class, i.e., the loss of information by summing up qualitatively different response patterns seems to be tolerable. Additional subscales with item-specific thresholds for dichotomization tended to be Rasch homogeneous: „affective" (three symptoms), „vegetative" (four symptoms), and „cognitive" (six symptoms). As usual, these findings should be cross-validated with another sample (in prep.). However, the results obtained so far for the dichotomized items may encourage making use of the subscale scores as an additional tool, and they may turn out to be helpful in improving predicition, for instance in regression/survival analysis.

Concerning the polytomous analyses, the models available are more sophisticated because of threshold parameter assumptions. In general, however, the subscales were found to be not Rasch homogeneous (cf. Kempf & Keller, 1992, for analyses of the Bouman & Kok-subscales with MIRA and LCA). The cognitive subscale proposed in this paper (with four categories within each item) does not seem to be Rasch homogenous as well. Furthermore, there are indications for a violation of the category ordering in several items for one class. The latter was apparent in other subscales as well (especially the symptoms „crying spells", „irritability"), and is already described by Frick, Rehm, and Thien (chapter 29).

Meanwhile, an official German version of the BDI (Hautzinger, Bailer, Worall & Keller, 1994) has been published. It will take some time until a sufficient amount of data is available for cross-validating our results with the new version. Since the wording of some items has been improved, it may be hoped that some of the shortcomings in scalability of categories will diminish. The new wording for the items, however, does not seem to affect the thresholds for dichotomization as used in this study. Replicability of the dichotomous analysis might thus be expected.

Acknowledgement

We appreciate the helpful comments made by the anonymous reviewer.

References

Beck, A.T., Steer, R.A. & Garbin, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77-100.

Bouman, T.K. & Kok, A.R. (1987). Homogeneity of Beck's Depression Inventory (BDI): Applying Rasch Analysis in Conceptual Exploration. *Acta Psychiatrica Scandinavica, 76,* 568-573.

Clark, D.C., von Ammon Cavanaugh, St. & Gibbons, R.D. (1983). The Core Symptoms of Depression in Medical and Psychiatric Patients. *Journal of Nervous and Mental Disease, 171,* 705-713.

Clark, D.C., Gibbons, R.D., Fawcett, J., Aagesen, C.A. & Sellers, D. (1985). Unbiased Criteria for Severity of Depression in Alcoholic Inpatients. *Journal of Nervous and Mental Disease, 173,* 482-487.

v. Davier, M. (1994). *WINMIRA. Program manual.* Kiel: IPN.

Hautzinger, M. (1991). Das Beck-Depressionsinventar (BDI) in der Klinik. *Nervenarzt, 62,* 689-696.

Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1994). *Beck-Depressions-Inventar (BDI).* Bern: Verlag Hans Huber.

Keller, F., Hautzinger, M., Wolfersdorf, M. & Steiner, B. (1991). Entlaß-Symptomatik als Prädiktor für Rückfall bei Depression: eine ereignisorientierte Auswertung. *Verhaltensmodifikation und Verhaltensmedizin, 12,* 186-200.

Keller, F. & Kempf, W. (1993). Stabilität und Validität latenter Klassen im Beck-Depressionsinventar (BDI). *Psychiatrische Praxis, 20* (Suppl.), 68-71.

Kempf, W. & Keller, F. (1992). *Psychometrische Analysen des Beck-Depressionsinventars (BDI).* Forschungsbericht PLK Weißenau, Ravensburg.

Richter, P. (1991). *Zur Konstruktvalidität des Beck-Depressionsinventars (BDI) bei der Erfassung depressiver Verläufe.* Regensburg: S. Roderer.

Richter, P., Werner, J. & Bastine, R. (1994). Psychometrische Eigenschaften des Beck-Depressionsinventars (BDI): Ein Überblick. *Zeitschrift für Klinische Psychologie, 23,* 3-19.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14,* 271-282.