

Chapter 6

Cross-validation of the Benzodiazepine Dependence Self-Report Questionnaire in Outpatient Benzodiazepine Users

C.C. Kan¹, M.H.M. Breteler², A.H.G.S. van der Ven³
and F.G. Zitman¹

¹Department of Psychiatry*, University of Nijmegen,
PO Box 9101, 6500 HB Nijmegen, The Netherlands

²Department of Clinical Psychology*, University of Nijmegen,
PO Box 9104, 6500 HE Nijmegen, The Netherlands

³Department of Mathematical Psychology, University of Nijmegen,
PO Box 9104, 6500 HE Nijmegen, The Netherlands

*University of Nijmegen Research Group on Addictive Behaviours (UNRAB)

Submitted for publication

ABSTRACT

The aim of this study was to cross-validate the Benzodiazepine Dependence Self-Report Questionnaire (Bendep-SRQ), which reflects the severity of benzodiazepine (BZD) dependence. The Bendep-SRQ, SCL-90, SCAN and ASI-R were administered to 102 general practice (GP) patients and 126 psychiatric outpatients who were using BZDs. The scalability and reliability of the Bendep-SRQ scales were re-assessed in terms of Rasch-homogeneity, subject discriminability, item discriminability and test-retest stability. Present and original Rasch item orders were compared to evaluate construct validity. A matrix of all measures was factor-analysed to assess concurrent and discriminant validity. The scalability of the Bendep-SRQ scales was confirmed. The reliability results were fairly good. Present and original Rasch item orders corresponded. The Bendep-SRQ scales and concurrent measures had high loadings on one factor, the discriminant measures on two other factors. In spite of some differences with respect to sociodemographic characteristics and pattern of BZD use, the cross-validation results agreed well with the results of the original study on the Bendep-SRQ. The Bendep-SRQ has presently acquired enough support of favourable and consistent results for clinical and scientific use.

INTRODUCTION

In 1995, Linsen et al. demonstrated the lack of consensus about the definition of benzodiazepine (BZD) dependence in a literature review.¹ At that time, a few self-report questionnaires had been described for the assessment of the BZD withdrawal syndrome.²⁻⁴ Since then, progress has been made with respect to more comprehensive diagnosis and psychometric severity assessment of benzodiazepine (BZD) dependence. Using the substance dependence sections of the semi-structured *Schedules for Clinical Assessments in Neuropsychiatry (SCAN)*,^{5,6} high prevalence rates of the DSM-III-R and ICD-10 BZD dependence diagnoses were found⁷ and subsets of the DSM-III-R and ICD-10 criteria were shown to constitute Rasch-homogeneous BZD dependence scales.⁸ Subsequently, the *Benzodiazepine Dependence Questionnaire (BDEPQ)* was developed by Baillie and Mattick to reflect the severity of BZD dependence.⁹ However, a multidimensional approach, suggested by the factor-analytical results, was not adopted in their report.

Recently, the *Benzodiazepine Dependence Self-Report Questionnaire (Bendep-SRQ)* was developed and its scalability, reliability and validity were assessed in a representative patient sample consisting of general practice (GP) patients, psychiatric outpatients and self-help patients.¹⁰ In the latter study a multidimensional approach was adopted. In order to construct proper scales, Rasch modelling was applied to the dimensions which were suggested by factor analyses. Promising results were found with respect to the scalability, reliability and validity of the four Bendep-SRQ scales which were outlined. Subsequently, a study with the same design was carried out on alcohol and drug dependent outpatients at Community-Based Addiction Centres (CBACs), which also yielded favourable results.¹¹

The aim of the present study was to cross-validate the Bendep-SRQ in new samples of GP

patients and psychiatric outpatients from other settings. The same scalability, reliability and validity assessments were repeated to investigate the generalizability of the promising results of the original study.¹⁰ Similar results were expected, which would further support the use of the Bendep-SRQ in clinical practice and scientific research.

METHOD

Settings and subjects

This study was conducted at five general practices and five psychiatric outpatient departments. To participate in the investigation the subjects had to meet the following inclusion criteria:

1) actual BZD use; 2) average frequency of BZD use of at least once a week; 3) age between 17 and 70 years; 4) ability to speak and read Dutch.

The patients who visited the general practices or psychiatric outpatient departments during the period of investigation, were screened according to these inclusion criteria. Eligible patients were asked to participate by a representative of the treatment team. Informed consent was obtained from 63% (102 out of the 162) of the GP patients and from 66% (126 out of the 191) of the psychiatric outpatients. The total sample of participants consisted of 228 subjects.

Study design

This study formed part of a larger project being conducted by the University of Nijmegen Research Group on Addictive Behaviours (UNRAB) in the Netherlands on the diagnosis of BZD dependence. The study population participated in two interviews, separated by three

weeks. During the first interview, sociodemographic data were collected, followed by administration of the Benzodiazepine Dependence-Self Report Questionnaire (Bendep-SRQ), the Benzodiazepine Dependence-Structured Diagnostic Interview (Bendep-SDI) and the Schedules for Clinical Assessments in Neuropsychiatry (SCAN).¹² The Bendep-SRQ and Bendep-SDI have been constructed by our own research group.¹⁰ The second interview, which was conducted by the same interviewer as the first, consisted of a re-administration of the Bendep-SRQ, followed by the Symptom Checklist-90 (SCL-90)¹³ and the Addiction Severity Index-Revised (ASI-R).¹⁴

Bendep-SRQ

The Bendep-SRQ was constructed at the Department of Psychiatry of the University Hospital Nijmegen, the Netherlands, with the aim of reflecting the severity of BZD dependence. The construction process of the Bendep-SRQ and its composition have been described previously.¹⁰ Analogously to this earlier study, the items of the Bendep-SRQ scales, which were originally 5-point rated, were dichotomized between the response options 2 (this is not true for me) and 3 (this is partly true, partly false for me) in order to apply Rasch analysis.

Scalability of the Bendep-SRQ scales

In the previous study by Kan et al. on GP patients, psychiatric outpatients and self-help patients, four Rasch homogeneous scales were extracted from the item pool of the Bendep-SRQ, which appeared to reflect 'Problematic Use', 'Preoccupation', 'Lack of Compliance' and 'Withdrawal'.¹⁰ Theoretical rationales were formulated to establish the construct validity of these scales. The scalability of these scales was confirmed in a second study on outpatients at

Community-Based Addiction Centres.¹¹ In the present study the Rasch analyses were repeated on the same scales in new samples of GP patients and psychiatric outpatients who were using BZDs.

Rasch analysis. By using the Bendep-SRQ scales, which are the sumscores of the dichotomized item responses, certain assumptions are implicitly made. These are specified in the Rasch model. To justify the use of the sumscores, these assumptions must be tested, which implies that the Rasch model should hold true. The assumptions from which the Rasch model can be derived¹⁵ and the required additive structure underlying the observed data, have been recapitulated in earlier reports.^{8,10} In essence, while the item responses depend on the respective underlying probabilities in a random way, the response probabilities themselves depend in a deterministic way on the subject and item scale values. According to the Rasch model, both subjects and items can be arrayed on a common unidimensional scale and the items have equal discriminative power (i.e. the property of equi-discriminability). Glas¹⁶ has developed two statistical tests for the dichotomous Rasch model, known as R1 and R2. The statistic R1 is especially sensitive to equi-discriminability, while the statistic R2 is sensitive to unidimensionality and local stochastic independence. If R1 is not significant at the 1% significance level ($P > 0.01$) the null hypothesis that all the items have equal discriminative power cannot be rejected and equi-discriminability can be assumed. Similarly, unidimensionality and local stochastic independence hold true when R2 is not significant ($P > 0.01$). Rasch-homogeneity is demonstrated if both statistics hold true, meaning that the sumscore across items is a sufficient statistic for the subject scale and that the sumscore across subjects is a sufficient statistic for the underlying item scale. To compute R1 and R2 the Rasch Scaling Program (RSP) was used.^{17,18}

Reliability

To evaluate the reliability of the Bendep-SRQ scales, the subject discriminability, item discriminability and test stability were assessed.

Subject discriminability (Internal Consistency). Subject discriminability implies that the subjects should differ systematically, i.e. the variation between subjects should be larger than the variation due to random error. The subject discriminability of the Bendep-SRQ scales was evaluated by means of the KR-20 coefficient. The size of KR-20 reflects the reliability of the scale, as the error variance of the estimator decreases if KR-20 increases.

Item discriminability. This should not be confused with the above-mentioned term equi-discriminability. It implies that the items should differ systematically, i.e. the variation between items should be larger than the variation due to random error. This was tested by Cochran's Q test.¹⁹ If the test result is significant, items can be considered to occupy distinct points on the scale. Additionally, analogous to the concept of reliability as described by Hoyt,²⁰ which is a measure of inter-subject discriminability, a measure of inter-item discriminability has recently been developed: the item discriminability coefficient (IDC).¹⁰ On the premises that the underlying item response model holds true, the IDC shows to what extent the differences between the items are systematic. The higher the IDC, the more powerful the predictions about the item scale will be.

Stability. To assess the test-retest reliability of the Bendep-SRQ scales, Pearson Product-Moment correlation coefficients were computed from the Bendep-SRQ data obtained at the first and the second interviews. The subjects who had discontinued their BZD use in the period between the interview sessions were excluded from the analysis.

Validity

The validity of the Bendep-SRQ scales was assessed in terms of construct, concurrent and discriminant validity.

Construct Validity. To establish the construct validity of the Bendep-SRQ scales theoretical rationales have been formulated¹⁰ to explain the specific item orders based on increasing Rasch scale values, reflecting increasing severity levels of the constructs. To comply with the postulated theoretical rationales, the estimates of the Rasch scale values in the present study should approximately replicate the specific item orders of the Bendep-SRQ scales in the former study.¹⁰ This would affirm the construct validity of the Bendep-SRQ scales.

Concurrent and Discriminant Validity. To investigate the concurrent and discriminant validity of the Bendep-SRQ we conducted Principal Axis factor analyses on the data matrix of the subjects who completed both interviews (n =225). The data matrix consisted of the sumscores of the Bendep-SRQ scales except for 'withdrawal' (in order to avoid the selection of patients with withdrawal experience only), the SCL-90 subscales, the ASI-R problem severity scores and either the Rasch-homogeneous ICD-10 or DSM-III-R BZD dependence scale. The latter two scales consisted of subsets of substance dependence items from the SCAN, as described in a separate paper.⁸ The concurrent validity of the Bendep-SRQ is supported when the Bendep-SRQ scales, the ICD-10 or DSM-III-R BZD dependence scale

and the ASI problem severity score for drug use (which includes BZD use) load substantially on a common factor, which can be interpreted as a BZD dependence factor. If the sumscores of the SCL-90 subscales and the remaining ASI problem severity scores load substantially on different factors, this supports the discriminant validity of the Bendep-SRQ scales.

RESULTS

Sociodemographic features and pattern of BZD use

Table 1 shows that women were overrepresented in both samples, especially in the GP sample. Most patients had a Dutch cultural background, were married, living with their partner and had been educated to primary level. Relatively more psychiatric outpatients than GP patients were receiving disability benefits. The 'mean daily BZD dose/defined daily BZD dose (MDD/DDD)' ratio was calculated to reflect the mean daily dose of all BZDs by one parameter; the 'DDD' is the defined daily dose recommended by the WHO. If more than one BZD was used, the average of the separate MDD/DDD values was used. On average, the GP patients were using a lower mean BZD dose than the recommended therapeutic level (MDD/DDD = 0.6), while the psychiatric outpatients were just exceeding it (MDD/DDD = 1.1). The GP sample contained the most chronic BZD users, shown by a mean duration of BZD use of 84 months and a P₇₅ value of 120 months.

Table 1. Sociodemographic characteristics and mean values for age, BZD dose and duration of BZD use

Variables	GP* patients (n=102)	Psychiatric Outpatients (n=126)
Sex		
male (%)^	23	42
female (%)	77	58
Mean age (years)±sd	53 ±10	47 ±10
Cultural Background (%)		
Dutch	99	98
Otherwise	1	2
Marital/social status (%)		
Single/never married	10	17
Engaged / steady relationship	6	6
Married	63	65
Divorced	11	9
Widowed	10	3
Living arrangement (%)		
Alone	25	20
With partner	72	73
Otherwise	3	7
Level of education (%)		
Primary level	60	47
Secondary level	28	33
Advanced level	13	21
Financial income (%)		
Profession	27	26
Unemployment benefit	8	5
Disability benefit	20	47
Pension	17	6
Partner's income	19	13
Otherwise	9	3
MDD/DDD#	.6	1.1
Quartiles	.1 - .4 - .9	.5 - .8 - 1.5
Mean duration of BZD~ use (months)	84	63
Quartiles	12 - 47 - 120	5 - 22 - 72

*GP: General Practice

^(%) : all percentages are given in rounded figures

#MDD/DDD : Mean Daily BZD Dose/Defined Daily BZD Dose

~BZD : if more than 1 BZD was being used, the duration was based on the BZD which had been used the longest

Table 2. Test results of Rasch analyses on Bendep-SRQ scales by means of RSP

Scale	i	R1	df	p	g	R2	df	p	n
Problematic Use	5	10.36	8	.24	3	10.86	8	.21	161
Preoccupation	5	9.55	4	.05	2	14.30	8	.07	170
Lack of Compliance	5	12.16	4	.02	2	13.51	8	.10	78
Withdrawal	5	9.45	4	.05	2	11.59	8	.17	106

RSP : Rasch Scaling Program²⁵

R1 : test statistic of Rasch analysis with regard to equi-discriminability²⁴

R2 : test statistic of Rasch analysis with regard to unidimensionality and local stochastic independence²⁴

i : number of items in the scale

df : degrees of freedom

p : p value

g : number of subgroups formed by Rasch analysis

n : number of subjects left in the analysis

Table 3. Reliability of the Bendep-SRQ scales in terms of subject discriminability, item discriminability and test stability

Parameter	Problematic Use			Preoccupation			Lack of Compliance			Withdrawal [#]		
	GP	PO	TS	GP	PO	TS	GP	PO	TS	GP	PO	TS
KR 20	.47	.62	.58	.76	.65	.72	.81	.72	.76	.82	.80	.82
CQ	72.30	49.61	113.06	125.65	172.87	296.45	39.69	35.37	70.77	8.33	9.60	17.75
p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.08	.048	.001
IDC	.95	.93	.97	.98	.98	.99	.90	.89	.95	.54	.58	.74
TRT [§]	.64	.67	.66	.81	.70	.76	.79	.77	.78	.77	.70	.75
p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

GP : General Practice patients (n =102)

PO : Psychiatric Outpatients (n =126)

TS : Total Sample (n =228)

[#] : Respondents who had never reduced or discontinued BZD use were excluded, leaving: GP: n =88, PO: n =105, TS: n = 193

KR-20 : Kuder-Richardson-20 coefficient of internal consistency

CQ : Cochran's Q

IDC : Item Discriminability Coefficient

TRT : Test-Retest Correlation

[§] : Subjects who had discontinued BZD use before the retest session were excluded, leaving: GP: n =95, PO: n =119, TS: n =216

Scalability

As shown in Table 2, the R1 and R2 test results of the Rasch analyses on the Bendep-SRQ scales were all non-significant ($P>0.01$), which implies that the Rasch model was not rejected at all. Therefore, the former Bendep-SRQ Rasch scales¹⁰ were confirmed in the present GP and psychiatric outpatient samples.

Reliability

The subject discriminability, item discriminability and test stability results, shown in Table 3, all indicated good reliability with respect to the Bendep-SRQ scales 'Preoccupation' and 'Lack of Compliance'.

The 'Problematic Use' scale yielded a low and a moderate KR-20 value in the GP and the psychiatric outpatient sample, respectively. The TRT correlations were satisfactory and the IDC values were good.

With respect to the 'Withdrawal' scale, the moderate IDC values for both the GP and psychiatric outpatients contrasted with the KR-20 values and TRT correlations, which were very good. A non-significant Cochran's Q value in the GP patients indicated that the variation between the withdrawal items could have been due to random error in this group. However, when both groups were combined, good Cochran's Q and IDC results were found.

Construct Validity

The item-orders based on increasing scale value estimates yielded by the Rasch analyses, were identical to the item orders found in the former study on the Bendep-SRQ,¹⁰ except for the 'Withdrawal' scale. However, the differences between the present and former item-order found for this scale, did not exceed a range of twice the standard error of any of the scale

value estimates. Therefore, including this approximate replication, these findings further supports the good construct validity of the Bendep-SRQ scales.

Discriminant and Concurrent Validity

The results of the Principal Axis Factor Analyses with Varimax rotation are shown in Table 4. A three factor solution was recommended by the scree plot, showing a substantial decrease and a gradual further decline in the eigenvalues of additional factors. Of course, the interpretation of this factor solution given below is not necessarily the best one, but it appeared to be the most plausible. The highest loadings of the Bendep-SRQ scales, the Rasch homogeneous ICD-10 or DSM-III-R BZD dependence scale and the ASI-R problem severity score for drug use were observed on the second factor that was extracted, reflecting a dimension of BZD dependence. These findings suggested good concurrent validity of the Bendep-SRQ scales.

The distribution of the factor loadings of the remaining scales supported good discriminant validity of the Bendep-SRQ scales. The first factor appeared to be a dimension of the psychopathological status, as all the SCL-90 subscales and the ASI-R severity score on psychiatric problems showed the highest loadings on this factor. The highest loadings of the ASI-R severity scores on professional and social problems were found on the second factor, suggesting that they were primarily associated with the dependence dimension. The problem severity for physical health showed the highest loading on the third factor, together with a number of secondary loadings of some SCL-90 scales and ASI-R severity scores, which was interpreted as a dimension of the somatic status. In as far as the secondary loadings were

Table 4. Principal Axis Factor Analyses with Varimax Rotation on a matrix (n = 225) consisting of scale scores

Scale	Factor		
	I	II	III
Bendep-SRQ			
Problematic Use		.48(.47)	
Preoccupation	.30(.33)	.61(.57)	
Lack of Compliance		.44(.52)	
SCAN			
ICD-10* (DSM-III-R*) past year BZD dependence		.63(.59)	
SCL-90			
Anxiety	.87(.87)		
Agoraphobia	.67(.68)	.31	
Distrust and Interpersonal Sensitivity	.86(.85)		
Depression	.88(.88)		
Insufficiency in thinking and acting	.81(.81)		
Hostility	.71(.70)		
Sleeping problems	.46(.46)		.34(.35)
Somatization	.71(.72)		.36(.38)
Remaining items	.84(.85)		
ASI-R problem severity areas			
Physical			.69(.68)
Professional		.51(.52)	.44(.44)
Alcohol			
Drugs		.67(.61)	.34(.33)
Social	.35(.37)	.47(.44)	
Psychiatric	.63(.65)	.40(.36)	

NOTE. Loadings between parentheses refer to the analysis using the DSM-III-R instead of the ICD-10 scale. Factor loadings of smaller than .3 are not shown.

*Rasch-homogeneous subset of BZD dependence criteria¹⁷

greater than 0.3 (see Table 4), they all agreed well within the above-mentioned interpretation of the factors.

DISCUSSION

Following the same design, GP patients and psychiatric outpatients who were using BZDs were assessed in the present study in order to cross-validate the promising results found in the first study on the Bendep-SRQ.¹⁰ The response rates and the sociodemographic characteristics were similar, but of course there were some differences. More of the present GP patients were married, instead of being single, and were consequently living with their partner instead of alone and more of them had been educated to primary instead of advanced level. All these features were strikingly similar in both the psychiatric outpatient samples. A clear difference was encountered in the GP patient samples with respect to the mean MDD/DDD ratio, which was lower (0.6) than in the former study (0.9). The mean duration of BZD use was higher in the present psychiatric outpatient sample than in the former one (63 versus 40 months). These kinds of differences between the study populations emphasize once more the value of cross-validation in order to justify the generalization of results.

The above results provide further support for the scalability, reliability and validity of the Bendep-SRQ scales when they are applied to GP and psychiatric outpatient samples. Moreover, the scalability results yielded by the Rasch analyses confirmed the multidimensional scale structure which has been outlined previously.^{10,11}

In comparison with the former study,¹⁰ low and moderate KR-20 values of the 'Problematic Use' scale, in the GP patients and psychiatric outpatients respectively, provided

less support for the subject discriminability. The low KR-20 value of .47 in the GP patients was probably due to the homogeneous nature of this sample with respect to 'awareness of problematic BZD use'. This might be related to the lower mean MDD/DDD ratio in the present GP sample (0.6 as compared to 0.9 in the original study); lower-dose BZD use could have made the 'degree of awareness of problematic BZD use' a less reliable measure by narrowing the response set to lower values. However, the subject discriminability of the other scales was not affected, which supports the reliability of the Bendep-SRQ. The IDC values did not provide as much support for the item discriminability of the 'Withdrawal' scale as they did before. This may have been due to the smaller number of responding subjects, because those who had never reduced or discontinued their BZD use were excluded. In terms of test stability, the TRT correlations indicated satisfactory reliability of the 'Problematic Use' scale and good reliability of the other scales.

The Rasch analyses in the present sample reconfirmed the construct validity of the Bendep-SRQ scales analogously to the study on the properties of the Bendep-SRQ in CBAC outpatients.¹¹ Not only Rasch-homogeneity was confirmed, but also the original item orders described by theoretical rationales.¹⁰

The interpretation of the factor analytical results shown in Table 4 resembled the interpretation in the first study on the Bendep-SRQ.¹⁰ This supports the concurrent and discriminant validity. In both studies, the first and the second factors reflected the psychopathological status and BZD dependence respectively, while the last factor reflected the somatic status. However, three factors were found in the present study as opposed to four in the former one. The additional factor of psychosocial problem severity in the former study, indicated by the highest loadings of the ASI-R severity scores on alcohol, professional, social and psychiatric problems, was not extracted presently. This was probably due to poorer

quality of the ASI-R severity scores. Unlike the Bendep-SRQ, SCAN and SCL-90 scales, these scores have been shown to be proper scales by means of scalability assessment.

Moreover, doubts about the unidimensionality of these ASI-R severity scores were raised in the former study by a number of secondary factor loadings of greater than 0.3.

In addition to the former studies on the psychometric properties of the Bendep-SRQ in GP patients, psychiatric outpatients, self-help patients¹⁰ and CBAC outpatients,¹¹ the present study provides further support for the scalability, reliability and validity of the Bendep-SRQ. Therefore, the Bendep-SRQ appears to have gained enough support for clinical and scientific applications. A subsequent report will deal with the standardization of the Bendep-SRQ in order to make the interpretation of the sumscores more feasible. Further research should shed more light on the predictive validity of the Bendep-SRQ scales. Certain critical Bendep-SRQ scale scores or particular profiles of the scores might warrant more rigorous treatment interventions. The effects of certain dosage regimens or discontinuation programmes on the severity of BZD dependence could be evaluated using the results of Bendep-SRQ monitoring. Attention should also be paid to possible interacting effects of persistent psychosocial stressors and personality traits. The Bendep-SRQ is recommended to clinicians and researchers as it is a practical and objective instrument which might help them in their efforts to make a difference in the prevention and management of BZD dependence.

NOTE

The Bendep-SRQ can be obtained from the authors (C.Kan@czzopsy.azn.nl) and is also available for on-line administration on site <http://baserv.uci.kun.nl/~fzitman/Bendep-SRQ.html>.

ACKNOWLEDGMENT

We thank all the staff from the general practices in Nijmegen, Beuningen, Weurt and Elst and the psychiatric outpatient departments in Nijmegen, Apeldoorn, Deventer, Arnhem and Venlo for contributing to this study.

REFERENCES

1. Linsen SM, Breteler MHM, Zitman FG. Defining benzodiazepine dependence: the confusion persists. *European Psychiatry* 1995;10:306-311.
2. Merz WA, Ballmer U. Symptoms of the barbiturate/benzodiazepine withdrawal syndrome in healthy volunteers: standardized assessment by a newly developed self-rating scale. *J Psychoactive Drugs* 1983;15:71-84.
3. Busto UE, Sykora K, Sellers EM. A clinical scale to assess benzodiazepine withdrawal. *J Clin Psychopharmacol* 1989;9:412-416.
4. Tyrer P, Murphy S, Riley P. The Benzodiazepine Withdrawal Symptom Questionnaire. *J Affect Disord* 1990;19:53-61.
5. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN, Schedules for clinical assessment in neuropsychiatry. *Arch Gen Psychiatry* 1990;47:589-593.
6. Janca A, Ustun TB, Sartorius N. New versions of World Health Organization instruments for the assessment of mental disorders. *Acta Psychiatr Scand* 1994;90:73-83.
7. Kan CC, Breteler MHM, Zitman FG. High prevalence of benzodiazepine dependence in outpatient users, based on the DSM-III-R and ICD-10 criteria. *Acta Psychiatr Scand* 1997;96:85-93.
8. Kan CC, Breteler MHM, Van der Ven AHGS, Zitman FG. An evaluation of the DSM-III-R and ICD-10 benzodiazepine dependence criteria using Rasch modelling. *Addiction* 1998;93(3):349-359.
9. Baillie AJ, Mattick RP. The benzodiazepine dependence questionnaire: development, reliability and validity. *Br J Psychiatry* 1996;169(3):276-281.
10. Kan CC, Breteler MHM, Timmermans EAY, Van der Ven AHGS, Zitman FG. Scalability, reliability and validity of the Benzodiazepine Dependence Self-Report Questionnaire in outpatient benzodiazepine users. *Compr Psychiatry* 1999;40(4):283-291.
11. Kan CC, Breteler MHM, Van der Ven AHGS, Zitman FG. Assessment of benzodiazepine dependence in alcohol and drug dependent outpatients. (Submitted).
12. Giel R, Nienhuis FJ. [SCAN. Schedules for Clinical Assessment in Neuropsychiatry] SCAN. Vragenschema's voor de klinische beoordeling in de neuropsychiatrie. Geneve: World Health Organization: [division for Mental Health] divisie voor de Geestelijke Gezondheid, 1992.
13. Arrindell WA, Ettema H. Dimensionele structuur, betrouwbaarheid en validiteit van de Nederlandse bewerking van de Symptom Checklist (SCL-90); gegevens gebaseerd op een fobische en een 'normale' populatie. *Ned T v Psychologie* 1981;36:77-108.
14. Hendriks VM, Kaplan CD, Limbeek Jv, and Geerlings P. The addiction severity index: reliability and validity in a Dutch addict population. *J Subst Abuse Treat* 1989;6:133-141.
15. Fischer GH. Derivations of the Rasch Model. In: Fischer GH, Molenaar IW (eds): *Rasch models: Foundations, Recent Developments and Applications*. New York: Springer Verlag, 1995:15-38.
16. Glas CAW. The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika* 1988;53:525-546.
17. Glas CAW. RSP: Rasch Scaling Program. Groningen, The Netherlands: IEC ProGAMMA: 1993.

18. Glas CAW. RSP: Rasch Scaling Program User's Manual. Groningen, The Netherlands: IEC ProGAMMA: 1993.
19. Siegel S. The case of k related samples. Non-parametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956:161-165.
20. Hoyt C. Test reliability by analysis of variance. Psychometrika 1941;6:135-160.

