Chapter 7

Latent Trait Standardization of the Benzodiazepine Dependence Self-Report Questionnaire using the Rasch Scaling Model

C.C. Kan[1], A.H.G.S. van der Ven[2], M.H.M. Breteler[3]

and F.G. Zitman[1]

[1]Department of Psychiatry*, University of Nijmegen,

PO Box 9101, 6500 HB Nijmegen, The Netherlands

[2]Department of Mathematical Psychology, University of Nijmegen,

PO Box 9104, 6500 HE Nijmegen, The Netherlands

[3]Department of Clinical 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 the present study was to obtain standardized scores that correspond with the raw scores on the four Rasch scales of the Benzodiazepine Dependence-Self Report Questionnaire (Bendep-SRQ). The eligible normative group for standardization of the Bendep-SRQ scales consisted of 217 General Practice (GP) patients, all using benzodiazepines. Two standardization methods were used and compared: ‘classical standardization’ which transforms raw scores into standard scores on the unit normal distribution and ‘latent trait standardization’ which transforms raw scores into latent trait scores. The latter requires the Rasch model with the additional assumption of a normally distributed latent trait, which held true for the scales ‘Problematic Use’, ‘Lack of Compliance’ and ‘Withdrawal’, but not for ‘Preoccupation’. The observed ‘unequal item spacing’ on the ‘Preoccupation’ scale was hypothesized to induce a response tendency of ‘non-deviation’, causing a local violation of the assumption of a normally distributed latent trait. Nevertheless, comparison of the results of the two standardization methods revealed such a high degree of resemblance, that latent trait standardization could be used for ‘Preoccupation’ just as well as classical standardization.

The presented standard scores and corresponding percentile ranks make raw Bendep-SRQ scores clinically interpretable in relation to the normative GP sample. Incorporation of the Rasch scaling methodology into the development of the Bendep-SRQ marks the adoption of the item response theory in the field of applied test methodology. In this process, it appears that ‘equal item spacing’ has to be taken into account to prevent local violations of the Rasch model with the additional assumption of a normally distributed latent trait.

INTRODUCTION

Previous reports have outlined the development of the Benzodiazepine Dependence Self-Report Questionnaire (Bendep-SRQ), an instrument which to reflect the severity of benzodiazepine (BZD) dependence comprehensively.[1-3] So far, these studies have concentrated on the investigation of the psychometric properties of the Bendep-SRQ in outpatients who were using benzodiazepines (BZDs). Four Bendep-SRQ item sets consistently met the scalability requirements of the Rasch model and therefore constituted proper Rasch homogeneous scales. Corresponding with the specific Rasch item orders, theoretical rationales could be formulated to support the construct validity of these scales, which were condensed into the designations 'Problematic Use', 'Preoccupation', 'Lack of Compliance' and 'Withdrawal'.[1] In terms of subject discriminability, item discriminability and test-retest stability, the reliability of these scales appeared sufficiently good and their concurrent and discriminant validity was supported by factor-analytical results.[1-3]

As a result of these studies, the Bendep-SRQ is presently a self-report instrument which provides a reliable and valid profile of four raw scale scores with regard to the severity of BZD dependence. However, such raw scores on psychometric instruments are usually too arbitrary for meaningful interpretation in clinical practice. The values of the mean, standard deviation and possible range of scores depend to a large extent on the formulation and selection of items prior to test construction. Unless the mean, standard deviation and the shape of the score distribution are known, no proper interpretation can be attached to the original raw scores. To solve this problem, it is customary to standardize psychometric instruments with respect to a normative population and derive standardized scores or 'norms'. The aim of the present study was to derive such norms for the Bendep-SRQ scales, on the basis of a normative population of outpatient benzodiazepine (BZD) users. For this purpose, conversion tables were composed, which permit the transformation of raw scores into standard scores. In addition to the classical method of transformation, a new and more theoretically-orientated method of transformation was applied based on the Rasch scaling model.[4,5] The latter was justified, because the Rasch model has already been used to demonstrate the scalability of the four Bendep-SRQ scales. Below, this new method of transformation is referred to as ‘latent trait standardization’, as it is based on a special case of the Rasch model, i.e. the Rasch model with the extra assumption that the latent subject trait is normally distributed.

METHOD

Standardization

To be able to interpret results of psychological tests, it is necessary to compare the raw scores to standards obtained from larger groups of subjects. A psychological test is said to be standardized when transformed scores, so-called norms, are available, which are based on a reference group of acceptable size. The scores of subjects take on meaning in relation to this standard, c.q. normative group. This procedure of obtaining norms is known as standardization.

Normative sample

To standardize scales, they have to be administered to a sample of subjects which properly represents the population for which the test is intended. The aim of the present study was to derive norms referring to a general standard population of outpatient BZD users. From earlier studies on the Bendep-SRQ, three potential normative samples were available: General Practice (GP) patients, psychiatric outpatients and outpatients at community-based addiction centres (CBACs).[2] In the Netherlands, a general practitioner usually prescribes BZDs to alcohol and drug dependent outpatients and chronic psychiatric outpatients. As such, CBAC outpatients and many psychiatric outpatients are also represented in a GP population. Therefore, the GP sample was considered to be the eligible normative sample for the computation of norms for the Bendep-SRQ scales.

Settings and subjects

The data for this study were obtained from nine general practices. 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. Patients who visited the general practices during the period of investigation were screened according to these inclusion criteria. Eligible patients were asked to participate by the GP's assistant. Informed consent was obtained from 65% (217 out of 334) of the eligible GP patients.

Study design

The present 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 detection and diagnosis of BZD dependence. The study population participated in two interviews, described in detail in earlier reports,[1,3] in which sociodemographic data were gathered and several questionnaires and structured interviews were administered, including the Bendep-SRQ. The present study concentrated entirely on the Bendep-SRQ data acquired in the first interview. The sociodemographic characteristics of the GP samples have been reported previously.[3,6,7]

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.[1] 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.

Conventional Rasch analysis

By using the Bendep-SRQ scales, which are the sum scores of the dichotomized item responses, certain assumptions are implicitly made. These are specified in the Rasch model. To justify the use of the sum scores, these assumptions must be tested, which implies that the Rasch model should hold true. The assumptions from which the Rasch model can be derived[5] and the required additive structure underlying the observed data, have been recapitulated in previous reports.[1,6] 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). Glas[8] 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 sum score across items is a sufficient statistic for the subject scale and that the sum score across subjects is a sufficient statistic for the underlying item scale.

Preceding standardization of the raw Bendep-SRQ scale scores, the Rasch-homogeneity of the items in the Bendep-SRQ scales was re-assessed with respect to the present normative GP sample by testing the Rasch model as described above, i.e. without any additional assumptions. To compute R1 and R2 the Rasch Scaling Program (RSP) was used.[9,10]

Standardization using the Classical Method

The classical method of test standardization is a frequently used procedure, which consists of the transformation of the raw score into the normal form. Test norms are usually normal transformations of the original raw scores with arbitrarily selected means and standard deviations. To transform scores into the normal form, every percentile rank, which is the cumulative percentage corresponding with a point on the observed score distribution, is matched with a specific point on the baseline of the unit normal curve measured in standard deviation (SD) units from a mean of 0. For example, a percentile rank of 50 corresponds with the zero point. A table of areas under the unit normal curve readily provides the number of SD units, the so-called z-score, which corresponds with the percentile rank. For instance, a percentile rank of 60 is .25 SD units above the mean. Thus, in the classical method of test standardization, percentile ranks are the stepping-stones in establishing the correspondence between a set of points on the original score scale and points on a normal distribution of zero mean and unit SD. Usually, z-scores are linearly transformed (multiplication by a constant and subsequent addition of a constant) to change the SD and the mean in order to eliminate negative values.

Standardization using the Latent Trait

In the Rasch model for dichotomous items, the item response function is defined as the probability of the response as a function of a latent trait.[4,5] A special case of the Rasch model makes the additional assumption that the latent trait has a normal distribution. Such a latent trait distribution could be used for the standardization of a scale instead of the observed score distribution. Theoretically, this appears to be only justified when the Rasch model with a normally distributed latent trait holds true. Therefore, before applying the model for standardization purposes, it should first be tested whether this special case of the Rasch model holds true or not. For this purpose RSP computes the statistic R0, which is based on a comparison of the observed score distribution and the expected score distribution. R0 is especially sensitive to violation of the assumption of a normally distributed latent trait.[9,10]

However, the statistics R1 an R2 are then also computed on the basis of a normally distributed latent trait. When the Rasch model with the additional assumption of a normally distributed latent trait holds true, RSP also generates aposteriori estimates, the so-called latent trait scores, which reflect the correspondance between the latent trait and the sum scores. These latent trait scores are computed as the expectation of latent trait given the sum score.[11,12]

Comparison of the standardization methods

To compare the Rasch latent trait method with the classical method, the latter was also applied to the expected score distribution (i.e. the expected sum score distribution when the Rasch model with a normally distributed latent trait holds true) instead of the observed score distribution (described above in the section on the classical method). To evaluate the differences between the standardization methods, the z-scores of these three methods were plotted in a scatter diagram. This evaluation might also answer the question whether the latent trait scores obtained by Rasch analysis could still be acceptable for the standardization of a scale which does not entirely meet the requirements of the Rasch model with the additional assumption of a normally distributed latent trait. If a similar linear relationship is shown by the plots of the z-scores of all mentioned standardization methods, then Rasch latent trait standardization can be applied just as well as classical standardization.

RESULTS

Conventional Rasch analysis

Results of the Rasch analyses on the Bendep-SRQ scales are displayed in Table 1. The results were all non-significant ($P > 0.01$), which confirmed the 'goodness of fit' of the Rasch model without the additional assumption of a normally distributed latent trait for all four Bendep-SRQ scales in the present normative GP sample.

Classical standardization

Table 2 illustrates transformation of the raw sum scores of the 'Problematic Use' scale into the normal form. The actual sum scores correspond with the exact midpoints of the class intervals in the second column. The third column shows the frequencies of the test scores, while the fourth column shows the cumulative frequencies at the midpoints. For example, the cumulative frequency at the midpoint of the interval 1.5 to 2.5 is equal to the number of cases below the interval added to half the frequency within the interval, i.e. (62 +61) + (0.5 x 55) = 150.5. The fifth column shows the percentile ranks at the midpoints of the intervals, which are the corresponding cumulative percentage frequencies at the midpoints. The sixth column shows

Table 1. R1- and R2 statistics of the conventional Rasch model for the Bendep-SRQ scales

Scale	R1	df	p	R2	df	p
Problematic Use	7.03	8	0.534	13.63	8	0.092
Preoccupation	6.55	8	0.585	7.79	8	0.454
Lack of Compliance	10.49	4	0.033	11.36	8	0.182
Withdrawal	12.57	8	0.128	8.76	8	0.363

df: number of degrees of freedom

p: right tail probability of chi square

Table 2. Classical transformation of the observed sum scores (x) into normal standard deviation units (z) for 'Problematic Use' (N = 216)

Class interval	x (=midpoint)	Observed frequency*	Cumulative frequency*	Percentile rank* (=cumulative percentage)	z
4.5 - 5.5	5	6	213.0	98.61	2.20
3.5 - 4.5	4	16	202.0	93.52	1.52
2.5 - 3.5	3	16	186.0	86.11	1.09
1.5 - 2.5	2	55	150.5	69.68	0.52
0.5 - 1.5	1	61	92.5	42.82	-0.18
-0.5 - 0.5	0	62	31.0	14.35	-1.06

*: at midpoint

Table 3. Transformation of the Bendep-SRQ sum scores (x) into normal standard deviation units (z) for all four scales (of: observed frequency; pr: percentile rank at midpoint)

	Problematic Use			Preoccupation			Lack of Compliance			Withdrawal		
x	of	pr	z	of	pr	z	of	pr	z	of	pr	z
5	6	98.61	2.20	25	94.21	1.57	1	99.77	2.83	28	92.00	1.41
4	16	93.52	1.52	20	83.80	0.99	9	97.47	1.95	27	76.29	0.72
3	16	86.11	1.09	68	63.43	0.34	11	92.86	1.47	15	64.29	0.37
2	55	69.68	0.52	35	39.58	-0.26	16	86.64	1.11	14	56.00	0.15
1	61	42.82	-0.18	38	22.69	-0.75	33	75.35	0.69	27	44.29	-0.14
0	62	14.35	-1.06	30	6.94	-1.48	147	33.87	-0.42	64	18.29	-0.90

the corresponding points on the baseline of the unit normal curve in standard deviation units from a zero mean, the normalized standard scores corresponding with the midpoints of the original score intervals (z-scores). The percentage of the area of the unit normal curve below a standard score of 2.20 is 98.61, the percentage below a standard score of 1.52 is 93.52, and so on.

Table 3 shows transformation of the sum scores for all the Bendep-SRQ scales into the normal form.

Rasch latent trait standardization

RSP uses the method of Marginal Maximum Likelihood (MML) estimation to estimate the values of the mean and SD of the normally distributed latent trait. The results of the Rasch analyses with the additional assumption are displayed in Table 4.

The null hypothesis that the latent trait has a normal distribution was not rejected for ‘Problematic Use’, ‘Lack of Compliance’ and ‘Withdrawal’ ($P > 0.01$). Therefore, the Rasch latent trait standardization method can evidently be applied to these three scales. However, the question arose how to explain the significant result in the case of ‘Preoccupation’.

In Figure 1 the observed frequency distributions of the sum score as well as the expected frequency distributions are displayed. In the case of ‘Preoccupation’ relatively large discrepancies are salient between the observed frequencies and expected frequencies corresponding with scores 3 and 4. These discrepancies also made the largest contributions to R0.

Table 4. Statistics R0, R1 and R2 of the Rasch model with the additional assumption of a normally distributed latent trait for the Bendep-SRQ scales

Scale	R0	df	p	R1	Df	p	R2	df	P
Problematic Use	7.86	3	0.049	21.45	11	0.029	8.21	12	0.769
Preoccupation	15.60	3	0.001	18.29	11	0.075	9.25	12	0.682
Lack of Compliance	1.38	3	0.710	15.30	7	0.032	16.36	12	0.175
Withdrawal	7.47	3	0.058	30.59	11	0.001	12.54	12	0.404

df: number of degrees of freedom

p: right tail probability of chi square

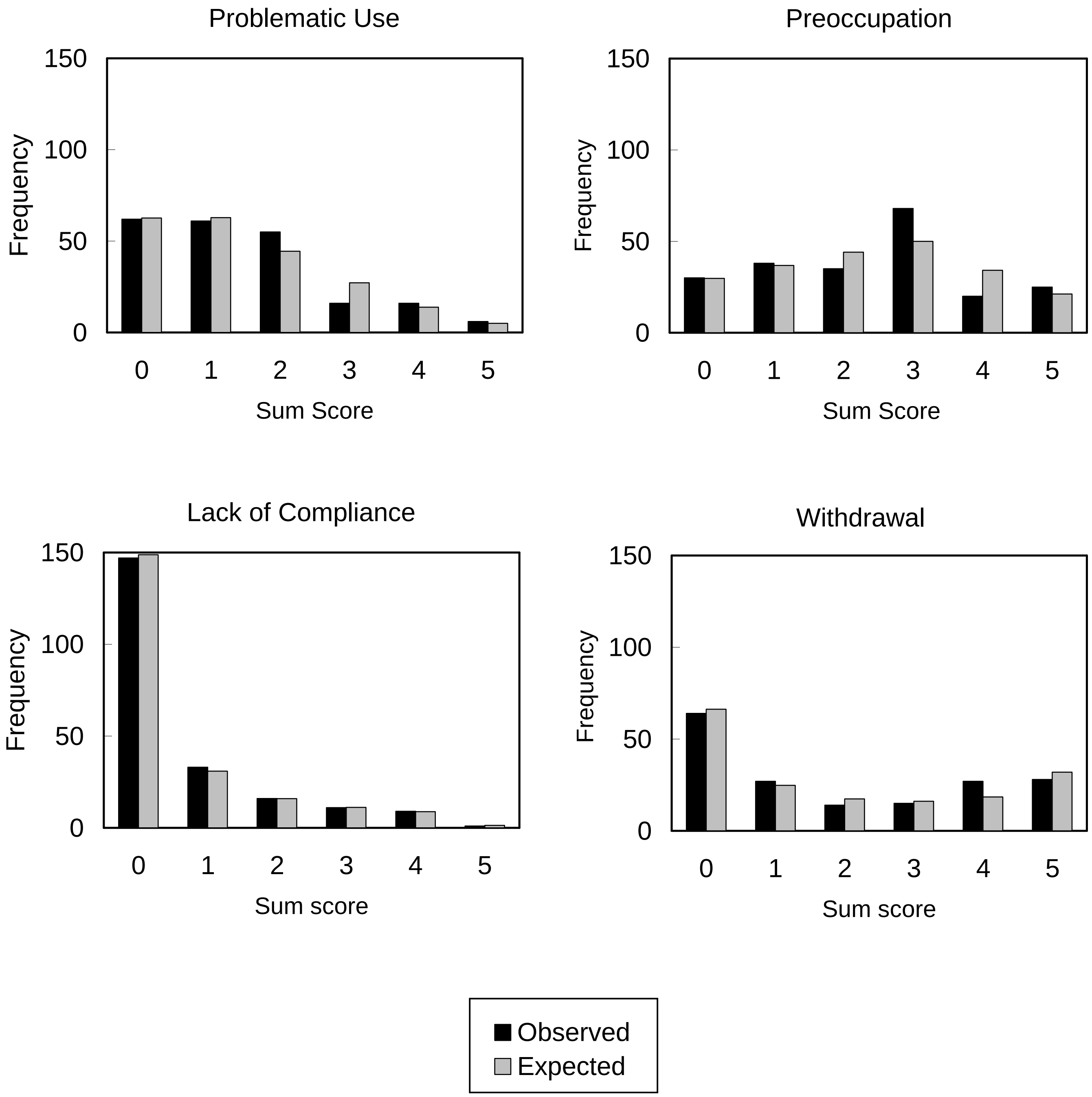


Fig 1. Frequency distributions of the observed and expected scores of the Bendep-SRQ scales.

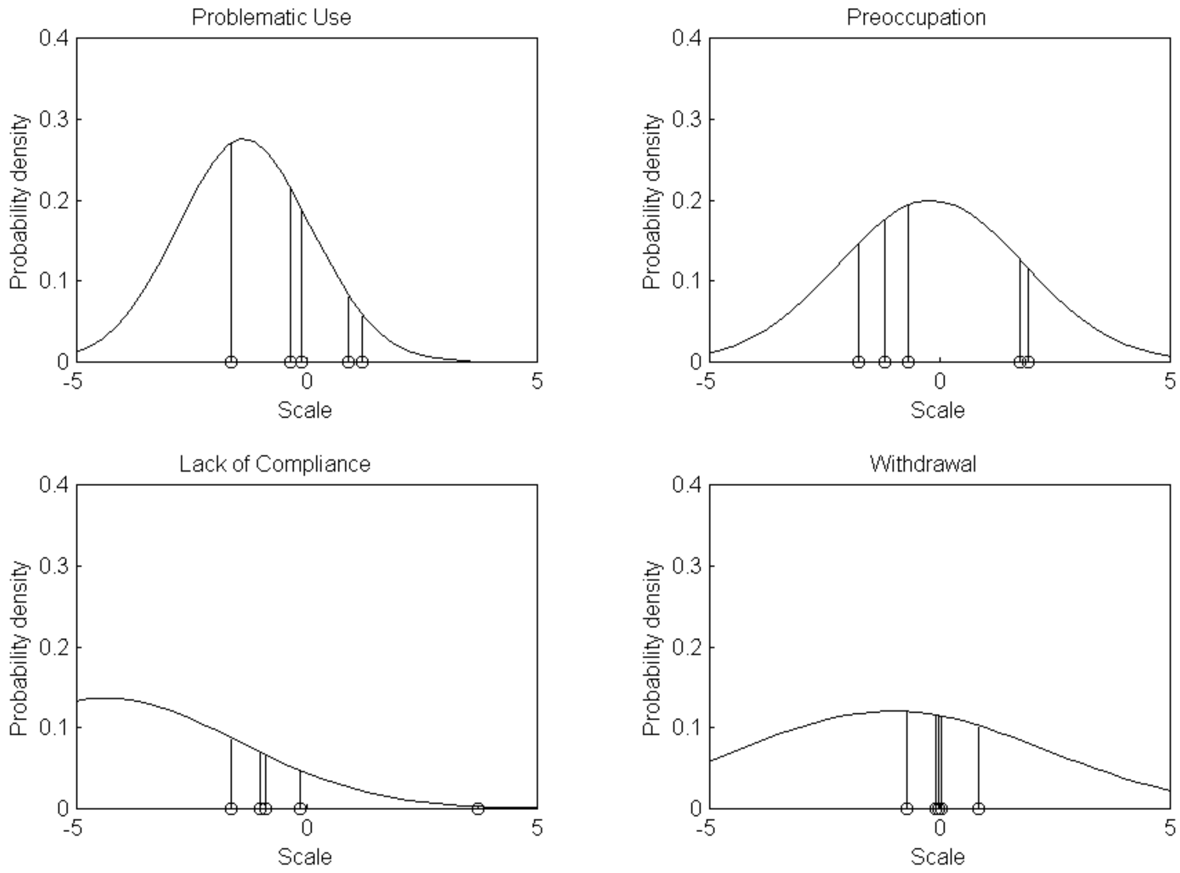


Fig. 2. Normal probability density and item scale values of the Bendep-SRQ scales yielded by Rasch analysis with the additional assumption of a normally distributed latent trait.

In Figure 2, which displays the item scale values for the Bendep-SRQ scales as yielded by Rasch analysis with the additional assumption of a normally distributed latent trait, it is salient that the 'Preoccupation' items are not equally spaced along the scale. In the case of 'Preoccupation', Figure 2 shows a gap between the third and fourth items. The significant R0 result for this scale can now be explained by assuming that subjects who are located in such a gap in the scale have a tendency 'not to deviate'; thus they respond positively to the previous item and negatively to the following item. This tendency of 'non-deviation' might explain why in the case of 'Preoccupation', the observed frequency of a sum score of 3 was higher than the corresponding expected frequency, while the observed frequency of a sum score of 4 was lower than the corresponding expected frequency (see Figure 2). According to this explanation, the observed number of subjects who responded positively to item 3 and had a sum score of 3 is expected to be larger than the expected number according to the model. Indeed, the observed number was found to be 68, whereas the expected number was equal to 49.98. Similarly, the observed number of subjects who responded positively to item 4 and had a sum score of 4, should be smaller than the expected number according to the model. In accordance with this expectation, the observed number was found to be 20, whereas the expected number was equal to 34.13. According to this explanation, the Rasch model is only violated locally, i.e. only by subjects with scale values within these gaps. However, the assumption of an underlying normally distributed latent trait has not necessarily been violated. Note that there is also a considerable gap between the fourth and fifth 'Lack of Compliance' items (see Figure 2). The number of subjects reflected by the area below the normal curve and between the two Rasch scale values adjacent to the gap can easily be calculated by converting the Rasch scale values into z-values on the unit normal curve with zero mean and a SD of 1. As the scale values of these items are -0.154 and 3.716, respectively, and

Table 5. Expected aposteriori estimates of the latent trait (lt), percentile ranks (pr) and standard scores(z) associated with the possible sum scores (x) of the Bendep-SRQ scales

	Problematic Use			Preoccupation			Lack of Compliance			Withdrawal		
x	lt	pr	z	lt	pr	z	lt	pr	z	lt	pr	z
5	1.40	97.13	1.90	2.80	93.33	1.50	2.90	99.38	2.50	3.44	91.32	1.36
4	0.63	91.48	1.37	1.50	80.26	0.85	1.02	96.86	1.86	1.44	77.37	0.75
3	-0.09	80.81	0.87	0.42	62.21	0.31	-0.29	92.09	1.41	0.38	66.68	0.43
2	-0.83	64.47	0.37	-0.59	42.11	-0.20	-1.41	84.87	1.03	-0.57	56.00	0.15
1	-1.65	42.11	-0.20	-1.62	23.61	-0.72	-2.79	70.92	0.55	-1.71	42.11	-0.20
0	-2.65	18.70	-0.89	-2.91	8.71	-1.36	-5.77	32.31	-0.46	-4.14	17.64	-0.93
Mean	-1.36		0	-0.19		0	-4.41		0	-1.05		0
SD	1.45		1	2.00		1	2.92		1	3.31		1

Severity designations for clinical interpretation of pr:

- **0 - 20 = ‘very low’**
- **20 - 40 = ‘low’**
- **40 - 60 = ‘moderate’**
- **60 - 80 = ‘high’**
- **80 - 100 = ‘very high’.**

Table 6. Classical transformation based on the expected score distribution for 'Problematic Use' (N = 216)

Class interval	x (=midpoint)	Expected frequency*	Cumulative frequency*	Percentile rank (Cumulative percentage*)	z
4.5 - 5.5	5	4.99	213.50	98.84	2.27
3.5 - 4.5	4	13.81	204.10	94.49	1.60
2.5 - 3.5	3	27.20	183.60	85.00	1.04
1.5 - 2.5	2	44.47	147.76	68.41	0.48
0.5 - 1.5	1	62.87	94.09	43.56	-0.16
-0.5 - 0.5	0	62.66	31.33	14.50	-1.06

*: at midpoint

the mean and SD of the latent subject distribution are -4.409 and 2.921, respectively, the proportion of subjects between these scale values is: P(-0.154 < s < 3.716) = P(1.457 < z < 2.782) = 0.99 - 0.92 = 0.07. This is only slightly more than five per cent. Due to this small proportion, the effect of the non-deviant response tendency was not strong enough to violate the Rasch model in this case. Likewise, the percentage of subjects between the scale values which belong to the third and fourth 'Preoccupation' items is equal to 43 (0.83-0.40). In order to compare these percentages, they were divided by their standardized scale distances, which gave 5 per cent for 'Lack of Compliance' and 35 per cent for 'Preoccupation'. These relative percentages show clearly why the effect of the non-deviant response was strong enough to reject R0 in the case of 'Preoccupation', but not in the case of 'Lack of Compliance'.

In Table 5 the latent trait score estimates (the expected aposteriori estimates of the latent trait scores) and the corresponding z-scores are shown for the Bendep-SRQ scales. The z-scores were computed with the corresponding mean and standard deviation of the latent trait distribution, which are given at the bottom of Table 5. As the Rasch model with the additional assumption of a normally distributed latent trait does not hold true for 'Preoccupation', the question arises as to whether it would still be justified to use the results shown in Table 5 to standardize these scales. In order to answer this question, these results were compared below with the standardization results using the classical method of standardization.

Table 7. Transformation of the sum scores (x) based on expected frequencies into normal standard deviation units (z) for the Bendep-SRQ scales (ef: expected frequency; pr: percentile rank at midpoint)

	Problematic Use			Preoccupation			Lack of Compliance			Withdrawal		
x	ef	pr	z	ef	pr	z	ef	pr	z	ef	pr	z
5	4.99	98.84	2.27	21.20	95.09	1.65	1.37	99.68	2.73	31.97	90.87	1.33
4	13.81	94.49	1.60	34.13	82.29	0.93	8.82	97.34	1.93	18.44	76.46	0.72
3	27.20	85.00	1.04	49.98	62.81	0.33	11.18	92.73	1.46	16.13	66.58	0.43
2	44.47	68.41	0.48	44.11	41.03	-0.23	15.91	86.49	1.10	17.43	57.00	0.18
1	62.87	43.56	-0.16	36.81	22.30	-0.76	30.93	75.70	0.70	24.76	44.94	-0.13
0	62.66	14.50	-1.06	29.77	6.89	-1.48	148.80	34.29	-0.40	66.27	18.93	-0.88

Comparison of standardization methods

While Table 2 shows the results of the transformation of the sum scores (x) based on the observed frequencies into normal standard deviation units (z) for ‘Problematic Use’, Table 6 shows the results of the transformation of the sum scores (x) based on the expected frequencies into normal standard deviation units (z) for ‘Problematic Use’. In Table 7 the results of the latter transformation are displayed for all the Bendep-SRQ scales.

If one compares the z-values obtained from the different methods of standardization which have been used, which are:

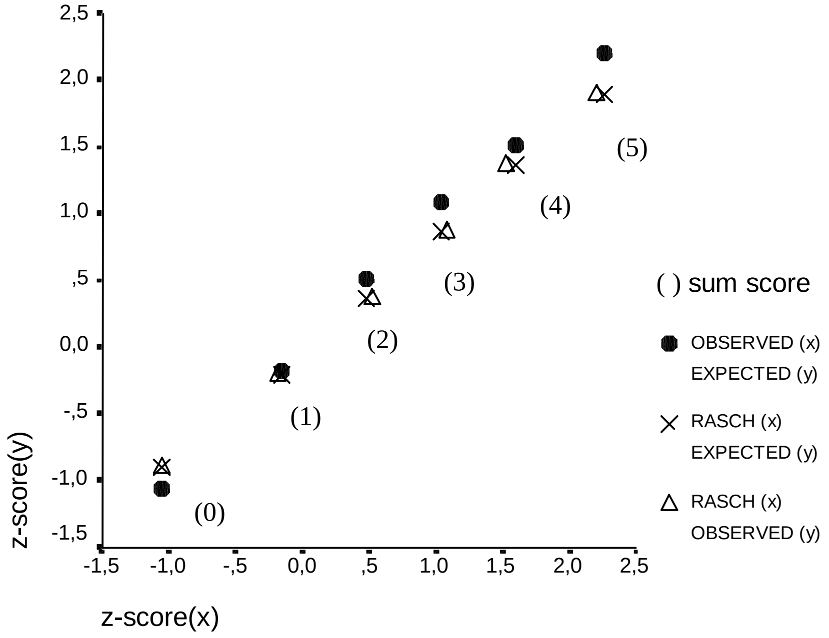
1. Classical standardization based on the observed frequencies (Table 3);
2. Rasch standardization based on the expected aposteriori estimates of the latent trait (Table 5);
3. Classical standardization based on the expected frequencies (Table 7),

then the following conclusions can be made:

A. The differences in z-values between the two classical methods (observed frequencies versus expected frequencies) are negligible. This is caused by the fact that the observed and expected frequencies do not differ very much.

B. The z-values obtained with the Rasch standardization are more conservative in comparison to the z-values obtained with the classical methods. For example, in the case of ‘Problematic Use’ , the z-values, which were obtained with the Rasch method ranged from -0.89 to +1.90, whereas the z-values, which were obtained with the classical methods ranged from -1.06 to +2.20 (observed frequencies) and from -1.06 to +2.27 (expected frequencies).

3a. Problematic Use



3b. Preoccupation

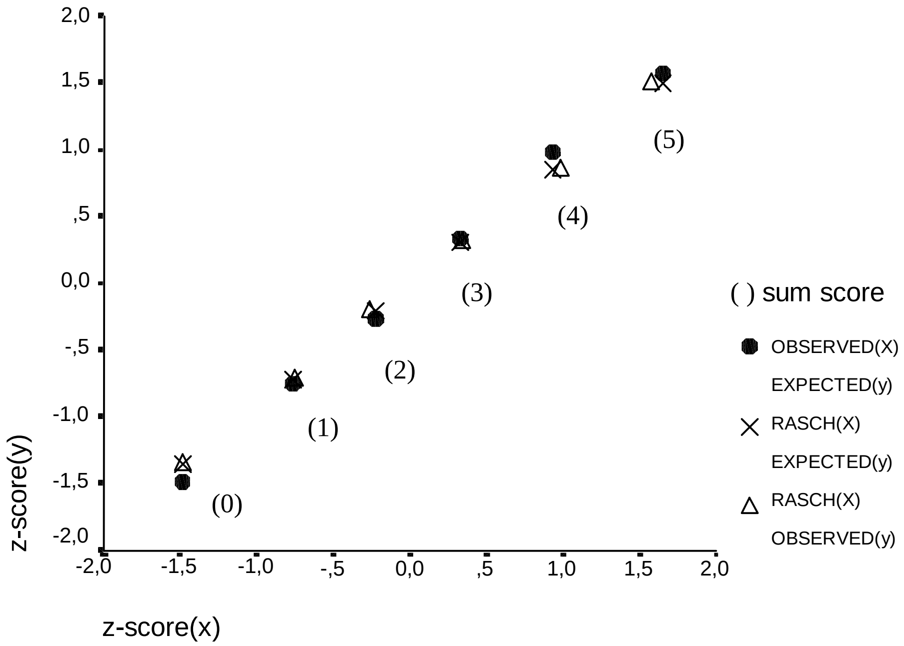


Fig. 3. Scatterplots of the z-scores of the three applied standardization methods on ‘Problematic Use’(3a) and ‘Preoccupation’(3b).

C. The z-values, obtained with all three methods, are approximately linearly related, as is clearly shown by the scatter plots in Figure 3. This was the case for ‘Preoccupation’ (Figure 3b) just as well as ‘Problematic Use’ (Figure 3a), despite the finding that the Rasch model with the additional assumption of a normally distributed latent trait did not hold true for ‘Preoccupation’. From these linear relationships between all three standardization methods, it follows that the Rasch latent trait method appeared to be just as acceptable as the classical method for the standardization of all the Bendep-SRQ scales, including ‘Preoccupation’. As Rasch latent trait standardization may, from a theoretical point of view, also be considered as the more elegant method, it was therefore decided to choose this method for the final standardization of the Bendep-SRQ scales.

DISCUSSION

The primary aim of the present study was to derive standard scores or norms that correspond with the raw sum scores of the Bendep-SRQ scales, in order to facilitate the clinical interpretation of the BZD dependence severity profile composed by these scores. For this purpose, the ‘latent trait standardization’ method was introduced by applying a special case of the Rasch scaling model to the Bendep-SRQ scales, i.e. the Rasch model with the additional assumption of a normally distributed latent trait. In contrast with the classical method of test standardization, latent trait standardization is theoretically backed-up by item response theory models, in which an assumption is made about the distribution of the latent trait in a normative population. This implies that standard scores can be derived which correspond with estimates of the latent subject trait instead of the observed scores. This is preferable, because the primary aim of a scale is to reflect the latent subject trait as accurately as possible.

Previously, conventional Rasch analyses have demonstrated the homogeneity of the items of the Bendep-SRQ scales,[1-3] which was confirmed in the present normative GP sample. The ‘goodness of fit’ of the Rasch model with the additional assumption of a normally distributed latent trait was demonstrated for the scales ‘Problematic Use’, ‘Lack of Compliance and ‘Withdrawal’, but rejected for ‘Preoccupation’. However, comparison of ‘latent trait standardization’ with ‘classical standardization’ revealed such consistent linear associations between the z-scores, that latent trait standardization also appeared to be suitable for the ‘Preoccupation’ scale. As a result, Table 5 in this paper presents standard scores and percentile ranks that correspond with the latent trait estimates of the Bendep-SRQ scales. These make sumsores observed in clinical practice immediately interpretable in relation to the normative GP sample. For example, if a patient has a raw score of 3 on ‘Problematic Use’, this corresponds with a latent trait score of -0.09, a z-score of 0.87 and a percentile rank of 80.81 (see Table 5). Likewise, the raw scores 2 and 4 correspond with the percentile ranks of 64.47 and 91.48, and so on. For each Bendep-SRQ scale, the percentile ranks given in Table 5 clearly outline the relative positions of subjects on a normal curve which reflects the latent subject trait. To guide clinical interpretation of the percentile ranks, corresponding designations of severity are stated in the legends of Table 5. Hence, the severity designations corresponding with the above-mentioned scores on ‘Problematic Use’ are ‘very high’, ‘high’ and ‘very high’ respectively. This interpretation makes it clear that a score of 1 reflects much greater severity of the latent traits of all the Bendep-SRQ scales than a score of 0. It also shows that the ‘Lack of Compliance’ scale is particularly discriminative with respect to the highest levels of its latent trait, because scores of 2 to 5 all fall into the ‘very high’ category.

While searching for the reason why the Rasch model with the additional assumption of a normally distributed latent trait did not hold true for one of the four Bendep-SRQ scales, it was hypothesized that ‘unequal item spacing’ on a Rasch scale can lead to local violation of the assumption of a normally distributed latent trait if the proportion of subjects in a relatively wide gap is relatively large. As an explanation for this phenomenon, a response tendency of ‘non-deviation’ was postulated, which means that a larger space between items causes bias in the item score of the items alongside the gap and consequently bias in the raw scores. Equal item spacing therefore appears to be a meaningful psychometric property for Rasch scales. With regard to the ‘Preoccupation’ scale, item spacing can be improved in the future by formulating new items which bridge the most salient gap. This would be facilitated by theoretical rationales that have previously been drawn up to reflect the specific item orders of the Rasch homogeneous Bendep-SRQ scales to support their construct validity.[1] According to the above-mentioned hypothesis, this bridging strategy by new items should cancel the local violation of the assumption of a normally distributed latent trait (R0: $P > 0.01$), if it succeeds in equalizing the item spacing.

Incorporation of the Rasch scaling methodology into the development of the Bendep-SRQ marks an important change, advocated by Duncan-Jones et al. in 1986,[13] i.e. the adoption of latent trait models in the field of applied test methodology. Beneath the properties reliability and validity, Rasch modelling has added a firm basis to the Bendep-SRQ scales by confirming their scalability[1-3] and, in the present study, by deriving standard scores and percentile ranks that correspond with estimates of the latent subject trait.

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, Malden, Beuningen, Weurt and Elst for contributing to this study.

REFERENCES

1. 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.

2. Kan CC, Breteler MHM, Van der Ven AHGS, Zitman FG. Assessment of benzodiazepine dependence in alcohol and drug dependent outpatients. (Submitted).

3. Kan CC, Breteler MHM, Van der Ven AHGS, Zitman FG. Cross-validation of the Benzodiazepine Dependence Self-Report Questionnaire in outpatient benzodiazepine users. (Submitted).

4. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960.

5. 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.

6. 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.

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. Glas CAW. The derivation of some tests for ther Rasch model from the multinomial distribution. Psychometrika 1988;53:525-546.

9. Glas CAW. RSP: Rasch Scaling Program. Groningen, The Netherlands: IEC ProGAMMA: 1993.

10. Glas CAW. RSP: Rasch Scaling Program User's Manual. Groningen, The Netherlands: IEC ProGAMMA: 1993.

11. Mislevy RJ. Estimating latent distributions. Psychometrika 1984;40:377-385.

12. Mislevy RJ. Bayes model estimation in item response models. Psychometrika 1986;51:171-195.

13. Duncan Jones P, Grayson DA, Moran PA. The utility of latent trait models in psychiatric epidemiology. Psychol Med 1986;16:391-405.