Rating catatonia in patients with chronic schizophrenia: Rasch analysis of the Bush–Francis Catatonia Rating Scale

ERIC WONG,¹ GABOR S. UNGVARI,² SIU-KAU LEUNG,³ WAI-KWONG TANG²

1 Centre for Epidemiology and Biostatistics, School of Public Health, Chinese University of Hong Kong, Hong Kong, China

2 Department of Psychiatry, Chinese University of Hong Kong, Hong Kong, China

3 Castle Peak Hospital, Hong Kong, China

Abstract

Background. Catatonic signs and symptoms are frequently observed in patients with chronic schizophrenia. Clinical surveys have suggested that the composition of catatonic syndrome occurring in chronic schizophrenia may be different from what is found in acute psychiatric disorders or medical conditions. Consequently, this patient population may need tailor-made rating instruments for catatonia. The aim of the present study was to examine the suitability and accuracy of using the Bush–Francis Catatonia Rating Scale (BFCRS) in chronic schizophrenia inpatients.

Method. The unidimensionality (optimal number of items; item fit), and the scoring scheme (the optimal number of scoring categories) of the BFCRS were determined in a random sample of 225 patients with chronic schizophrenia applying Rasch analysis. In addition, differential item functioning (DIF) analysis was also performed.

Results. The BFCRS proved to be unidimensional apart from three misfit and one marginally misfit items. The three misfit items were removed from the scale thereby constructing a revised version called BFCRS-R. Since the original BFCRS (BFCRS-O) showed no increase across items across steep gradients (poor endorsability of step calibrations), in BFCRS-R a binary scale ('absent' versus 'present' choices only) was constructed instead of the scoring scheme of 0–3.

The 20-item BFCRS-R showed improved psychometric properties in that it had a higher item separation index than BFCRS-O. BFCRS-R mean logit was closer to zero indicating that the items on the scale and the subjects were better matched than in BFCRS-O. DIF analysis showed that certain items of both versions of BFCRS were influenced by the presence of negative symptoms.

Conclusion. BFCRS-R is shorter and simpler than the original version and having better psychometric properties seems to be better suited for identifying and quantifying catatonia in chronic psychotic patients. Copyright © 2007 John Wiley & Son, Ltd.

Key words: catatonia, chronic schizophrenia, Bush–Francis Catatonia Rating Scale, Rasch analysis

Introduction

Catatonia has staged a remarkable comeback over the past 15–20 years (Fink and Taylor, 2003; Caroff et al., 2004). One of the major problems holding back further research on catatonia has been the fluidity of the boundaries of the concept. There is no consensus

what symptoms and signs constitute a catatonic syndrome or if whether there are, or how many distinct syndromes of catatonia (Ungvari and Carroll, 2004). This explains why the existing catatonia rating instruments differ so much from each other in terms of the number and definitions of individual symptoms

(Rosebush et al., 1990; Lund et al., 1991; Bush et al., 1996a, 1996b; Northoff et al., 1999; Braunig et al., 2000). This fact provides the rationale for continuing to re-examine the reliability and validity of catatonia rating scales.

The Bush–Francis Catatonia Rating Scale (BFCRS) was the first instrument constructed for the systematic, standardized and quantifiable examination of catatonia using operationally defined signs and symptoms (Bush et al., 1996a). The BFCRS consists of 23 items accompanied by specific instructions to standardize the examination. Seventeen items are scored on a 0–3 scale while the remaining six are rated as either absent ('0') or present ('3'). Selection of the items was based on a review of the classical and contemporary literature. The inter-rater reliability of BFCRS was established on 28 acutely ill patients presenting with catatonic syndrome that arose in the context of various psychiatric disorders. With the lack of biological indicators for catatonia, the BFCRS was validated against existing major studies regarded as clinical gold standard (Bush et al., 1996a). Catatonia ratings with the BFCRS predicted treatment response to lorazepam with high probability (Bush et al., 1996b).

When the architects of BFCRS applied it to a cohort of selected 42 chronic patients the composition of catatonia was found essentially the same as in acute patients (Bush et al., 1997). We had, however, different experience with chronic patients using the BFCRS. In a random sample of 225 chronically ill schizophrenia patients we found the distribution of catatonic symptoms quite different from that of Bush et al.'s acute and chronic patients; there were more mannerisms, stereotypy, posturing and grimacing and less stupor, mutism, and vegetative disturbances in our sample (Ungvari et al., 2005).

The division between acute and chronic catatonia was further evidenced by their treatment response. While 75–90% of acutely ill catatonia patients respond to benzodiazepines (Fink and Taylor, 2003), in a placebo-controlled, crossover study schizophrenia patients with longstanding (>5 years), prominent catatonic features were entirely resistant to a 6-week course of lorazepam (Ungvari et al., 1999).

As the BFCRS has been, and is likely to be used in several research projects to identify and measure catatonia, another look at this scale from a different angle is clearly warranted. Specifically, there is a pressing

need to examine the feasibility of applying the BFCRS in diagnostically different patient populations, to test its accuracy and efficiency, i.e. to determine the optimal structure of the scale that provides the maximum amount of information about the presence and severity of catatonia occurring in various diagnostic groups.

Item response theory (IRT), also called as latent trait models, is increasingly used to assess the properties of various health outcome measures including rating scales and questionnaires (Garratt, 2003). The common feature of IRT models is their ability to calibrate items of an assessment tool by their location on the latent trait; in our case, catatonic signs/symptoms along the underlying catatonia construct. The advantage of using IRT models is that they allow the selection of the most informative items.

One of the IRT models, Rasch analysis (Wright and Mok, 2000) has been applied in psychiatry, among others, for the assessment of mood (Aggen et al., 2005), alcoholism (Kahler et al., 2004) and psychosis (Graves and Weinstein, 2004).

In Rasch analysis, unidimensionality and item fit of a scale means that a single underlying construct is present which is measured by items that form the particular scale; for the BFCRS, if this scale is unidimensional, catatonic symptoms constituting the syndrome of catatonia must be accounted for by a single construct. A computer-assisted search (MEDLINE, EMBASE, PSYCHLIT) targeting catatonia found no papers that examined the unidimensionality and item fit of the BFCRS.

IRT stipulates that given the same trait level (e.g. severity of a symptom), the probability of affirming an item of the rating scale should be the same between groups (Scheuneman, 1979); in other words, the item should not have significant differential item functioning (DIF) for any subgroup of patients. DIF analysis has been used to examine the effect of culture (e.g. rural–urban subgroups) on various mood scales (Azocar et al., 2001). Further details of the DIF analysis are discussed in the Method section.

The aim of the present study was to examine the feasibility of using the BFCRS in patients with persistent catatonic signs and symptoms by applying Rasch analysis. Specifically, we set out to determine the unidimensionality, item fit, and DIF of the BFCRS for basic demographic and clinical variables in a sample of 225 Chinese patients with chronic schizophrenia.

Method

Subjects

The selection of subjects and details of the clinical assessment were described in previous papers (Leung et al., 2003; Ungvari et al., 2005). Briefly, 225 chronically ill patients with DSM-IV diagnosis of schizophrenia were randomly selected from eight of the 14 adult wards of a large psychiatric institution in Hong Kong. There were 160 men and 65 women in the sample; the mean age, age of onset and length of illness were 41.7 ± 7.4 , 21.3 ± 6.4 and 20.4 ± 7.5 years, respectively. The duration of current admission was 99.4 ± 80.9 months. The mean dose of antipsychotic drugs in chlorpromazine equivalent was 812 ± 593 mg/day.

The study protocol was approved by the Clinical Research Ethics Committee of the New Territories North Hospital Cluster. All subjects signed a written consent form.

Assessment of catatonia

Different aspects of the subjects' clinical condition were evaluated independently by three experienced psychiatrists using well-established rating scales (for details see Leung et al., 2003; Ungvari et al., 2005). In this paper we focus only on ratings of catatonia.

The mean number of catatonic signs/symptoms and the mean BFCRS score in the sample were 3.3 ± 2.7 (range: 0–12) and 4.4 ± 4.0 (range: 0–21), respectively. The most common catatonic symptoms were mannerisms (35.1%), grimacing (30.7%), stereotypy (24.4%), impulsivity (22.7%), mutism (21.3%), posturing (20.4%) and automatic obedience (20.0%; Ungvari et al., 2005).

Statistical analysis

The BFCRS was assessed for its unidimensionality, item fit, redundancy and DIF with the partial credit model, one of the Rasch models (Andiel, 1995; Wright and Mok, 2000) using the Winsteps computer program, Version 3.04 (MESA Press, Chicago, IL, USA), which implements an unconditional maximum likelihood procedure (Wright and Panchapakesan, 1969). Since the item difficulty of only one construct can be measured in Rasch analysis, therefore unidimensionality is the basic prerequisite of this method. The Rasch model assumes that ratings of a set of items can be explained by the characteristics of the examinee (patient) and those of the items, in this case the catatonic symptoms.

Patients' and item scores are used to 'calibrate' items on a logit scale. A logit is the natural log-odds of the level of difficulty of a particular item in relation to all other items in the scale. Items at one end of the scale are 'easier to detect' (i.e. more frequent) while items at the other end are more 'difficult to detect', that is, rare. The difficulty of individual items is determined by the frequency of endorsement, i.e. the presence and severity of a particular item. In the present analysis, items with a negative (–) calibration were more 'difficult', i.e. indicated more severe catatonia. Rasch analysis also constructs a hierarchy of the examinees; in this case, the hierarchy was ordered by the patients' severity of catatonia.

As mentioned, unidimensionality refers to the single underlying construct measured by items that form a scale. The adequacy of the fit of each item to the Rasch model is assessed by the information weighted mean-square residual goodness-of-fit statistics [information weighted fit (INFIT) and outlier-sensitive fit (OUTFIT)] (Wright and Masters, 1982) that are measures providing information about the scores given to items around the same difficulty endorsement level as the patient's characteristics. An acceptable range for an INFIT/OUTFIT value is between 0.7 and 1.3 (Wright and Linacre, 1994). An INFIT/OUTFIT value of less than 0.7 indicates that the item does not provide information beyond that was given by the rest of the items on the scale. This can occur when there are several similar or highly correlated items, or, when one item is dependent on another. An INFIT/OUTFIT value of greater than 1.3 indicates that the item does not define the same construct as do the rest of the items; it is either a poorly constructed or understood item, or it is ambiguously defined. Items with poor fit statistics should be considered for removal from the rating instrument.

DIF analysis (Holland and Wainer, 1993) was performed to determine the variability of item scores across subgroup of patients ascertained by sex, age, age of onset, length of illness, antipsychotic medication and the Brief Psychiatric Rating Scale (BPRS; Woerner et al., 1988) and Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1983) scores. Briefly, for each patient's score on each item the standardized residual of the observed score from what was predicted by the model was calculated. Then the Rumm software divided the subjects' logits into three groups of identical size according to the severity of catatonia. Thus each person was classified according to one of class intervals

giving a set of residuals suitable for a two-way analysis of variance (ANOVA) design. Thus, the statistical test used for detecting DIF is an ANOVA of the person-item deviation residuals with patient-related factors (e.g. age, sex and length of illness) and class intervals (i.e. the group according to the severity of catatonia) as factors.

For the DIF analysis, two classes were arbitrarily defined for age (≤ 45 and ≥ 45 years; 153 versus 72 subjects), age of onset (≤ 18 and ≥ 19 years; 83 versus 142 subjects), length of illness (≤ 20 and ≥ 21 years; 111 versus 114 subjects), antipsychotic medication [≤ 600 mg chlorpromazine equivalent (CPZeq) and ≥ 601 mg CPZeq; 109 versus 116 subjects], BPRS (sum scores of ≤ 25 and ≥ 26 ; 161 versus 64 subjects) and SANS (sum scores of ≤ 60 and ≥ 61 ; 102 versus 123 subjects). Two types of DIF can be identified: uniform and non-uniform DIF. With the former, there is a constant difference between groups in the probability of affirming an item (or category) across the trait (ANOVA main effect), and with the latter the difference varies across the trait (ANOVA interaction effect).

Results

The calibrated person and item difficulty map of the original BFCRS (BFCRS-O) is depicted in Figure 1. It shows the distribution of measures of the severity of the illness in terms of the signs and symptoms of catatonia and the item difficulty for the 23-item BFCRS-O on the same linear continuum based on the partial credit model. On the left of the vertical line are patients; more severe catatonic patients are on the top and less severe ones are on the bottom. Items are placed on the right side of the vertical line; more difficult items on the top and less difficult ones on the bottom. The mean logit score for the sample was -2.27 ± 0.52 indicating that the 23 items (mean logit 0.00 ± 0.23) on average were more likely to be rated with a higher than a lower score. The logit measures and the INFIT/OUTFIT statistics of individual items of the BFCRS-O are presented in Table 1. Eighteen (78.3%) of the 23 items had INFIT/OUTFIT statistics between 0.7–1.3 indicating that the scale was unidimensional, that is, these items contributed to a single underlying construct. The misfit items of #11 ('Rigidity'), #15 ('Impulsivity'), #20 ('Grasp reflex') and #23 ('Autonomic abnormality') were not closely related to this single construct, and item #3 ('Mutism') was redundant. The person and item separation indices were 0.74 and 2.97, respectively.

Twenty-one (91.3%) of the 23 items were scored '3' (the highest category of the BFCRS) less than 10 times or not at all. The average measures and the step calibration did not advance monotonically across the categories indicating that the items did not have optimal number of categories. (Statistics are not presented here but are available from the principal author upon request).

To correct the psychometric deficiencies of the BFCRS-O a revised version of the scale (BFCRS-R) was created with only 'present' and 'absent' categories for each item. Three misfit items were deleted: #11 ('Rigidity'), #20 ('Grasp reflex') and #23 ('Autonomic abnormality'); item #15 ('Impulsivity') which was marginally misfit (OUTFIT = 1.33) was retained in BFCRS-R. Applying BFCRS-R to the same sample, the mean logit score for the sample was -1.71 ± 0.85 , indicating that the 20 items (mean logit of items = 0.00 ± 0.87) on average were easier to endorse with a positive response by the rater compared to BFCRS-O. The person-item map of BFCRS-R is shown in Figure 2. The mean logit score for the sample of BFCRS-R is closer to the mean logit of items (= 0) when comparing with the corresponding figure with BFCRS-O. This means that items and subjects are better matched with BFCRS-R. The person and item indices were 0.71 and 3.56, respectively.

None of the items of BFCRS-R had a significant DIF for sex, age, age of onset, length of illness, antipsychotic medication and the BPRS and SANS scores, except for items 2 ('Excitement'), 3 ('Immobility'), 4 ('Mutism'), 5 ('Posturing') and 14 ('Withdrawal') which all displayed uniform DIF for SANS; subjects showing these items tended to have significantly higher SANS. Item 14 also displayed non-uniform DIF for age of onset; subjects who were withdrawn had significantly higher age of onset (Table 2). (As results of the DIF analysis were nearly identical for both versions of the scale, only the figures for BFCRS-R are shown in Table 2.)

Discussion

There has been growing interest in catatonia over the past decade (Fink and Taylor, 2003; Caroff et al., 2004) resulting in an ever increasing number of publications on the topic. Yet, the concept is still ill-defined and consequently a number of signs and symptoms diverse in their presentation and possibly etiology are subsumed under catatonia, mostly due to tradition. Different catatonia rating scales include varying number of items

Figure 1. Item-person map of the BFCRS-O. Each '#' represents four patients. M denotes mean logit; S denotes one standard deviation from the mean logit; T denotes two standard deviations from the mean logit.

Table 1. Logit measures, and INFIT/OUTFIT statistics for items of BFCRS-O and BFCRS-R

Item	BFCRS-R			BFCRS-O	
	Logit measure (error)	INFIT MNSQ	OUTFIT MNSQ	INFIT MNSQ	OUTFIT MNSQ
23 Autonomic abnormality	3.45 (1.01)	1.02	6.26	—	—
20 Grasp reflex	1.80 (0.46)	1.05	3.87	—	—
18 Gegenhalten	1.61 (0.42)	0.98	0.72	1.01	0.84
10 Verbigeration	1.06 (0.33)	1.01	0.86	1.01	0.88
13 Waxy flexibility	0.95 (0.32)	0.97	0.87	0.97	0.87
11 Rigidity	0.77 (0.30)	1.01	1.53	—	—
7 Echopraxia/echolalia	0.53 (0.27)	1.01	1.05	1.00	1.17
19 Ambitendency	0.53 (0.27)	1.00	0.99	0.99	1.00
12 Negativism	0.27 (0.25)	0.96	0.89	0.98	0.92
22 Combateness	−0.06 (0.22)	1.04	0.93	1.04	0.95
1 Excitement	−0.24 (0.21)	1.02	0.94	1.02	0.95
21 Perseveration	−0.24 (0.21)	1.00	0.95	1.00	1.00
14 Withdrawal	−0.53 (0.19)	0.92	0.86	0.93	0.88
2 Immobility/stupor	−0.57 (0.19)	0.95	0.83	0.97	0.85
17 Mitgehen	−0.60 (0.19)	0.99	0.97	0.99	0.99
4 Staring	−0.71 (0.19)	0.98	1.01	0.99	1.03
16 Automatic obedience	−0.81 (0.18)	0.96	0.85	0.95	0.87
5 Posturing/catalepsy	−0.84 (0.18)	0.93	0.92	0.94	0.95
3 Mutism	−0.91 (0.18)	0.86	0.75	0.87	0.77
15 Impulsivity	−1.00 (0.18)	1.16	1.33	1.17	1.36
8 Stereotypy	−1.13 (0.17)	1.07	1.08	1.07	1.10
6 Grimacing	−1.53 (0.17)	1.06	1.04	1.07	1.05
9 Mannerisms	−1.80 (0.16)	1.06	1.04	1.08	1.07

Note: MNSQ represents mean square; logit of smaller magnitude represents increasingly difficult item or more severe catatonia.

ranging from 10 (Rosebush et al., 1990) to 40 (Northoff et al., 1999).

To date, no sub-syndromes of catatonia have been clearly delineated. In view of the variety of neuropsychiatric and general medical conditions that can manifest with catatonia (Fink and Taylor, 2003) it is expected that a generic catatonia rating instrument, of which BFCRS is an example, would be, to some degree, insufficient to capture catatonia in specific patient populations and consequently would need modification. To this end we tested the BFCRS in a sample of patients with chronic schizophrenia where catatonic symptoms are frequently encountered. Classical authors (Bleuler, 1911/1950; Kraepelin, 1913/1919) found that up to 30% of patients with chronic schizophrenia displayed prominent catatonic features. In the few, but not all, modern

studies that systematically screened chronic psychotic population for catatonic symptoms, they were observed even higher proportion ranging from approximately 67% (Owens et al., 1982) to 100% (Rogers, 1985).

Using Rasch analysis, the original BFCRS (BFCRS-O) had four misfit items that did not measure the same construct than the rest of the scale so they were removed. A further problem with BFCRS-O was that the endorsability (step calibrations) did not increase across individual items with steep gradients meaning that scores on the BFCRS (0–3) did not represent distinct categories. Accordingly the scale was modified (BFCRS-R); three of the misfit items were removed together with the graded scoring system.

The results suggest that from a psychometric point of view the binary BFCRS-R is more suitable for

Figure 2. Item-person map of the BFCRS-R. Each '#' represents five patients. M denotes mean logit; S denotes one standard deviation from the mean logit; T denotes two standard deviations from the mean logit.

Table 2. Differential Item Functioning (DIF) analysis of the BFCRS-R ($n = 97$)

Item	BPRS		SANS		Sex		Age		Age of onset		Antipsychotic medication		Length of illness	
	U	NU	U	NU	U	NU	U	NU	U	NU	U	NU	U	NU
1 Excitement	0.3650	0.7380	0.1508	0.1369	0.9006	0.1361	0.1042	0.0232	0.3700	0.1703	0.1040	0.0704	0.0404	0.0641
2 Immobility/stupor	0.3244	0.3686	0.0001*	0.4278	0.1814	0.8923	0.4453	0.2599	0.3814	0.6100	0.8032	0.6163	0.8476	0.9034
3 Mutism	0.3602	0.1041	<0.0001*	0.0063	0.8080	0.9216	0.0395	0.1406	0.6581	0.5280	0.9919	0.1210	0.6745	0.5904
4 Staring	0.8752	0.0890	0.0011*	0.2118	0.0756	0.9007	0.0409	0.2035	0.7533	0.3001	0.0832	0.9584	0.7938	0.5827
5 Posturing/catalepsy	0.9103	0.3228	0.0011*	0.2363	0.4692	0.4354	0.1030	0.3078	0.2897	0.3892	0.4341	0.4228	0.3823	0.5716
6 Grimacing	0.2734	0.0532	0.0074	0.7590	0.6553	0.4469	0.4813	0.8395	0.1900	0.4979	0.6502	0.6954	0.9400	0.3000
7 Echopraxia/echolalia	0.8473	0.6105	0.6209	0.1597	0.2046	0.3912	0.0296	0.7827	0.4298	0.9380	0.7923	0.1633	0.9385	0.3782
8 Stereotypy	0.5699	0.9047	0.1723	0.2909	0.5182	0.5537	0.0725	0.0523	0.4288	0.0483	0.7747	0.7189	0.0540	0.7396
9 Mannerisms	0.1930	0.8217	0.0891	0.6921	0.2415	0.1541	0.7534	0.4254	0.3627	0.5355	0.7386	0.4938	0.9506	0.0148
10 Verbigeration	0.1085	0.0159	0.6007	0.1195	0.2999	0.1500	0.6218	0.4631	0.9785	0.9870	0.5989	0.9435	0.9640	0.0238
11 Negativism	0.0784	0.8294	0.0809	0.4433	0.2636	0.9573	0.5759	0.3444	0.7975	0.7696	0.0166	0.5162	0.9424	0.1083
12 Waxy flexibility	0.3907	0.6501	0.2216	0.0705	0.5634	0.3280	0.4706	0.1192	0.6949	0.7793	0.6862	0.3491	0.4885	0.2188
13 Withdrawal	0.0439	0.1656	0.0006*	0.0915	0.1645	0.8162	0.2057	0.1001	0.2898	<0.0001*	0.9858	0.9613	0.9424	0.9054
14 Impulsivity	0.3176	0.2356	0.0514	0.3615	0.8136	0.7252	0.0365	0.1127	0.0758	0.5220	0.9103	0.4209	0.9665	0.1653
15 Automatic obedience	0.1579	0.6977	0.0623	0.0770	0.4278	0.2092	0.0340	0.6204	0.8261	0.0812	0.6589	0.6877	0.7083	0.8838
16 Mirgehen	0.8066	0.8242	0.3106	0.8213	0.1094	0.3482	0.0570	0.5517	0.8635	0.2481	0.5110	0.3879	0.972	0.7373
17 Gegenhalten	0.3658	0.4228	0.8038	0.5157	0.1318	0.4507	0.5375	0.8996	0.4676	0.0271	0.2625	0.7627	0.7733	0.4892
18 Ambitendency	0.8394	0.2478	0.7585	0.1267	0.2632	0.2167	0.1476	0.8108	0.0364	0.9630	0.2484	0.4799	0.1611	0.7304
19 Perseveration	0.0202	0.0461	0.2372	0.2784	0.7569	0.7206	0.7170	0.8930	0.5417	0.3206	0.7913	0.5674	0.5012	0.8507
20 Combativeness	0.3474	0.8957	0.1564	0.5506	0.6069	0.4679	0.0256	0.1323	0.7098	0.1419	0.5556	0.7937	0.3082	0.7057

Note: because of multiple tests were performed, the level of significance of 0.05 was adjusted by Bonferroni correction to 0.0025 (0.05/20). U represents uniform DIF; NU represents non-uniform DIF.

*Significant at 0.0025 level.

measuring catatonia in patients with chronic schizophrenia. BFCRS-R has a higher item separation index and a mean logit closer to zero indicating a better fit between the items and the subjects' characteristics.

The significant correlation found between the SANS and items 'withdrawal', 'stupor', 'mutism' and 'posturing' of both the BFCRS-O/R indicates that catatonic and negative symptoms share certain features; whether this is the result of conceptual overlap or shared etiology of the two symptom domain warrants further investigation.

In conclusion, if the BFCRS-R was reduced to 20 items and had only 'present' and 'absent' categories for each item the psychometric figures of the scale improved compared to BFCRS-O. BFCRS-R would be more advantageous for clinical practice and research because it is shorter and simpler and faster in scoring catatonia in chronic patients. If these results were replicated BFCRS-R could be used for rating catatonia in chronic psychiatric patients.

References

- Aggen SH, Neale MC, Kendler KS. DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *Psychol Med*, 2005; 35: 475–487.
- Andiel C. Rasch analysis: a description of the model and related issues. *Can J Rehabil* 1995; 9: 17–25.
- Andreasen NC. The Scale for Assessment of Negative Symptoms (SANS). Iowa City, IA: University of Iowa, 1983.
- Azocar F, Arean P, Miranda J, Munoz RF. 2001. Differential item functioning in a Spanish translation of the Beck Depression Inventory *J Clin Psychol* 57: 355–365.
- Beuler E. *Dementia Praecox or the Group of Schizophrenias*, trans. Zinkin J. New York: International University Press, 1911/1950.
- Braunig P, Kruger S, Shugar G, Hoffler J, Borner I. The catatonia rating scale I: development, reliability and use. *Compr Psychiatry* 2000; 41: 147–158.
- Bush G, Fink M, Petrides G, Dowling F, Francis A. Catatonia. I. Rating scale and standardized examination. *Acta Psychiatrica Scandinavica* 1996a; 93: 129–136.
- Bush G, Fink M, Petrides G, Dowling F, Francis A. Catatonia II. Treatment with lorazepam and electroconvulsive therapy. *Acta Psychiatrica Scandinavica* 1996b, 93: 137–143.
- Bush G, Petrides G, Francis A. Catatonia and other motor syndromes in a chronically hospitalized psychiatric population. *Schizophrenia Res* 1997; 27: 83–92.
- Caroff SN, Mann SC, Francis A, Fricchione GL (eds). *Catatonia*. Washington, DC: American Psychiatric Publishing Inc., 2004.
- Fink M, Taylor MA. *Catatonia. A Clinician's Guide to Diagnosis and Treatment*. Cambridge: Cambridge University Press, 2003.
- Garrett AM. In collaboration with the United Kingdom Back Pain Exercise and Manipulation Trial. Rasch analysis of the Roland disability questionnaire. *Spine* 2003; 28: 79–84.
- Graves RE, Weinstein S. A Rasch analysis of three of the Wisconsin Scales of Psychosis Proneness: measurement of schizotypy. *J Appl. Meas* 2004; 5: 160–171.
- Holland PW, Wainer H (eds). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum, 1993.
- Kahler CW, Strong DR, Read JP, Palfai TP, Wood MD. Mapping the continuum of alcohol problems on college students: a Rasch model analysis. *Psychol Addict Behav* 2004; 18: 322–333.
- Kraepelin E. *Dementia Praecox and Paraphrenia*, trans. Barclay RM. Edinburgh: Livingstone, 1913/1919.
- Leung SK, Ungvari GS, Ng FS, Cheung HK, Leung T. Tardive dyskinesia in Chinese inpatients with chronic schizophrenia. *Progr. Neuro-Psychopharm. Biol. Psychiat.* 2003; 27: 1029–1035.
- Lund CE, Mortimer AM, Rogers D, McKenna PJ. Motor, volitional and behavioural disorders in schizophrenia. I. Assessment using the Modified Rogers Scale. *Brit J Psychiatry* 1991; 158: 323–327.
- Northoff G, Koch A, Wenke J, et al. Catatonia as a psychomotor syndrome: a rating scale and extrapyramidal motor symptoms. *Mov Disord* 1999; 14: 404–416.
- Owens DGC, Johnstone EC, Frith CD. Spontaneous involuntary disorders of movement – their prevalence, severity, and distribution in chronic schizophrenics with and without treatment with neuroleptics. *Arch Gen Psychiatry* 1982; 39: 452–461.
- Rogers D. The motor disorders of severe psychiatric illness: a conflict of paradigms. *Br J Psychiatry* 1985; 147: 221–232.
- Rosebush PI, Hildebrand AM, Furlong BG, Mazurek MF. Catatonic syndrome in a general psychiatric inpatient population: frequency, clinical presentation, and response to lorazepam. *J Clin Psychiatry* 1990; 51: 357–362.
- Scheuneman JD. A method of assessing bias in test items. *J Educ Meas* 1979; 16: 143–152.
- Ungvari GS, Carroll BT. Nosology. In Caroff SN, Mann SC, Francis A, Fricchione GL (eds) *Catatonia*, pp. 33–43. Washington, DC: American Psychiatric Publishing Inc., 2004.
- Ungvari GS, Chiu HFK, Chow LY, Lau BST, Tang WK. Lorazepam for chronic catatonia: a randomized, double-blind, placebo-controlled cross-over study. *Psychopharmacology* 1999; 142: 393–398.
- Ungvari GS, Leung SK, Ng FS, Cheung HK, Leung T. Schizophrenia with prominent catatonic features ('catatonic schizophrenia'). I. Demographic and clinical correlates in the chronic phase. *Progr Neuro-Psychopharm Biol Psychiat* 2005; 29: 27–38.

- Woerner MG, Mannuzza S, Kane JM. 1988. Anchoring the BPRS: an aid to improved reliability. *Psychopharmacol Bull* 1988; 24: 112–117.
- Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Measurement Transactions* 1994; 8: 370.
- Wright BD, Masters GN. *Rating Scale Analysis*. Chicago, IL: Mesa Press, 1982.
- Wright BD, Mok M. Rasch models overview. *J Appl Meas* 2000; 1: 83–106.
- Wright BD, Panchapakesan N. A procedure for sample-free item analysis. *Educ Psychol Meas* 1969; 29: 23–48.
- Correspondence: Gabor S. Ungvari, Shatin Hospital, Shatin, N.T. Hong Kong SAR, China. Telephone +852-2636-7570; Fax +852-2647-5211. Email: gsungvari@cuhk.edu.hk*