Validity and reliability of the Polish version of the Positive and Negative Syndrome Scale (PANSS)

MAŁGORZATA RZEWUSKA, Institute of Psychiatry and Neurology, Warsaw, Poland

ABSTRACT *This study presents the results of a training course on using the Positive and Negative Syndrome Scale (PANSS) for the assessment of mental status. The agreements between ratings produced by 418 physicians (psychiatrists) were evaluated and compared with standard ratings. In this way it was possible to supplement some of the listed symptoms with comments, which may help to increase agreement between results obtained through raters' assessments and standard ratings.*

Key words: PANSS, reliability, Polish version

Introduction

The Positive and Negative Syndrome Scale (PANSS), a tool used in assessing the psychopathology of schizophrenia, was first published in 1987. Its authors (Kay, Opler and Lindenmayer, 1987) based their work on the 18-item BPRS scale, which has been widely used in clinical studies since 1962 (Overall and Gorham, 1988) and on the PRS (Psychopathology Rating Scale), which covered negative symptoms and those related to the impairment of social functions (Kay, 1990). The PANSS allows for the assessment of 30 items listed in three subscales covering positive, negative and general symptoms. The scale has become a standard tool for evaluating the dynamics of abnormalities in clinical drug trials.

The original scale, as well as its versions in other languages (Von Knorrig and Lindström, 1992; Lindström et al., 1994; Müller et al., 2000), have been evaluated for reliability and validity (agreement of ratings).

In Poland, translations of the rating scale have been in use for many years, mostly in international multi-centre trials of new anti-psychotic agents. Physicians using the PANSS scale should undertake training, which can be done using original video recordings containing patient assessments and interpreting comments by the scale's authors. Having participated in many international training sessions together with other Polish psychiatrists, I noticed that our assessment results regarding several items of the scale constantly

differed from standard ratings. This was true of items concerning statements reflecting subjective experiences (depression, anxiety) and certain items where the score depended on delusional experiencing of the self (for example being concerned with one's health). Polish psychiatric tradition is closely linked with German psychopathology, where symptoms are evaluated on the basis of factors other than pure observation (and the literal content of statements). Thus, Polish physicians tend to base their judgements on evaluations of relationships between statements and the manner in which they are expressed rather than on their pure content. These observations prompted me to undertake the evaluation of the PANSS with the participation of a large number of psychiatrists.

The aim of the study was to evaluate the agreement of mental state ratings of the schizophrenic patient performed with the Polish version of the PANSS and to compare them with original rating results of the authors of the scale (the standard rating). The possible identification of causes for discrepancies would allow for the formulation of comments, which would highlight the intent of the scale's authors.

Method

I performed the evaluation using a video recording of the examination of one patient. The patient, with chronic schizophrenia (PANSS 126 points), was examined by Kay (the author of the PANSS). The film

contained the full examination of the patient, the ratings of the authors of the scale (the standard rating) and the justification for these ratings.

The video cassette was supplied by Janssen-Cilag Poland and it was the property of this company, which agreed to the use of the cassette for training purposes. The film was preceded by an expert's presentation of individual items of the scale according to definitions supplied in the manual. The examination was carried out in English but the film contained subtitles in Polish. Table 1 presents an overview of the structure of the sessions.

Ratings were performed by 418 psychiatrists (in 22 groups numbering 17–25 members). Each assessment was followed by Kay's comments and an analysis of the agreement of ratings proposed by individual participants and a discussion of the differences between participants' results and the standard ratings.

Inter-rater agreement was evaluated according to Kay's (1991) method.

Ratings were regarded as being in agreement if their score did not differ from the standard (original) rating by more than 1 point. The number of correct ratings was recorded for subscales (positive, negative, general) and for the overall scale. According to the literature, 70% agreement is 'acceptable' whereas 90% agreement is 'high'. If 80% of participants show agreement of 80%, the training result is considered satisfactory (Kay et al., 1988; Von Knorring and Lindström, 1992; Lindström et al., 1994; Lambert, 1996). I used Cochran's Q-test to evaluate agreement with the standard rating statistically, setting the significance level at $p = 0.05$.

I also calculated Cohen's concordance ratio (Kw kappa) to evaluate correct ratings with an error of ± 1 point.

I considered accidental agreement of ratings using the ICC rate, which measures the statistical significance of agreement of assessments (Landis and Koch, 1977). Ratings were classified as concordant if the ICC rate exceeded 0.90 and as satisfactory if the ICC rate was in the range of 0.40 to 0.50.

Table 1. Structure of sessions

Introduction	10'
PANSS video	40–45'
Completion of rating forms	30–40'
Video – Kay's estimation and discussion of rating	60–90'

Results

Analysis of inter-rater agreement

Mean ratings in the positive, negative and general subscales were acceptable. Rating agreement levels (± 1) were as follows: 76% for the positive subscale, 81.2% for the negative subscale, 73.8% for the general symptom subscale and 76% for the overall scale.

The level of inter-rater agreement was good with respect to 23 symptoms (77%). Over 90% of physicians assessed nine symptoms identically, whereas over 80% proposed similar ratings for 14 symptoms. Table 2 presents the findings for agreement in assessments.

A slightly smaller group (77% to 79%) proposed concordant ratings of a further five symptoms. The lowest level of agreement was observed in assessment of the severity of mannerisms (71%) and concern for own somatic health (65%). Such results may suggest difficulties in the evaluation of these symptoms or varying interpretation of items. The high variability observed in scores makes the latter explanation more likely.

Analysis of agreement between observer and standard rating

The agreement between ratings proposed by group members and the standard ratings was lower than inter-rater agreement. There was a high level of inter-rater agreement (90%) in regard to 20% of symptoms, but just under 50% of symptoms were correctly assessed by less than 80% of the raters. The respective levels of agreement with standard ratings for individual symptoms are presented in Table 3 (concordant ratings) and Table 4 (discordant ratings). Table 5 presents the items where inter-rater agreement was higher than rater-standard agreement, and Table 6 presents rating scale items that were over- or under-rated.

Discussion

This study demonstrates that mental state ratings performed by Polish psychiatrists show a number of differences when compared with ratings by authors of the PANSS. Inter-rater differences are much smaller and they meet criteria for a sufficient level of agreement.

It was impossible to compare all the results of this study to the results of other authors. The other studies show only the ratings of the subscales of the PANSS and the global assessment.

Based on a comparison of ratings and assessments given during discussions that followed the training

Table 2. Inter-rater agreement [Inter-rater agreement (N = 418)]

Symptom	A. Very high agreement of ratings	
	Percentage of inter-rater agreement	Kappa (rating distribution)
N2 Emotional withdrawal	94	0.94
P2 Conceptual disorganization	93	0.93
G9 Unusual thought content	93	0.29
N6 Lack of spontaneity and flow of conversation	92	0.91
P1 Delusions	91	0.17
G4 Tension	91	0.21
G13 Disturbance of volition	91	0.13
N4 Passive/apathetic social withdrawal	90	0.35
N5 Difficulty win abstract thinking	90	0.99
Symptom	B. High agreement of ratings	
	Percentage of rater agreement	Kappa (rating distribution)
P6 Suspiciousness/persecution	89	0.98
N7 Stereotyped thinking	89	0.14
G2 Anxiety	89	0.14
G10 Disorientation	89	0.97
P3 Hallucinatory behaviour	87	0.98
G11 Poor attention	87	0.99
G12 Lack of judgment and insight	85	0.26
N3 Poor rapport	84	0.98
P7 Hostility	83	0.44
G14 Poor impulse control	82	0.95
P5 Grandiosity	81	0.21
G15 Preoccupation	81	0.95
G16 Active social avoidance	80	0.96
G3 Feelings of guilt	80	0.10
Symptom	C. Lowest agreement of ratings	
	Percentage of rater agreement	Kappa (rating distribution)
N1 Blunted affect	79	0.94
G6 Depression	79	0.30
P4 Excitement	78	0.43
G8 Uncooperativeness	78	0.11
G7 Motor retardation	77	0.37
G5 Mannerisms and posturing	71	0.18
G1 Somatic concern	65	0.96

sessions, one may propose the following as explanations of the discrepancies between ratings.

Delusions (P1)

The severity of delusions was rated higher by 68% of participants. This follows from the premise that delusional associations and thinking are crucial for determining the patient's behaviour and relationship with his or her environment. Traditionally, Polish psy-

chiatry does not perceive systematization of delusions as signifying higher severity of disturbance. Kay in the recorded presentation explained that the high rating of delusions for this patient was related not only to the amount of delusional content but its systematization. Therefore, it seems appropriate to supplement this item in the Polish version with a comment that the rater must consider the systematization of content, which is assumed to be a measure of severity.

Table 3. High rater-standard agreement

Symptoms	Percentage of inter-rater agreement (A)	Percentage of rater-standard agreement (B)	A–B	Kappa (A v/s B)	ICC (distribution)
P2 Conceptual disorganization	93	93	0	0.93	0.93
P3 Hallucinatory behaviour	87	87	0	0.98	0.98
P6 Suspiciousness/persecution	89	89	0	0.98	0.97
N1 Blunted affect	79	79	0	0.94	0.94
N2 Emotional withdrawal	94	94	0	0.94	0.94
N3 Poor rapport	84	84	0	0.98	0.98
N5 Difficulty in abstract thinking	90	90	0	0.99	0.98
N6 Lack of spontaneity and flow of conversation	92	92	0	0.91	0.91
G9 Unusual thought content	93	90	3	0.29	0.30
G10 Disorientation	89	84	5	0.97	0.97
G11 Poor attention	87	87	0	0.99	0.99
G14 Poor impulse control	82	91	1	0.95	0.95
G15 Preoccupation	81	81	0	0.95	0.95
G4 Tension	91	91	0	0.21	0.28
G16 Active social avoidance	80	80	0	0.98	0.98

Table 4. Low rater-standard agreement

Symptoms	Percentage of inter-rater agreement A	Percentage of rater-standard agreement B	A–B	Kappa	ICC
P7 Hostility	83	72	11	0.44	0.46
P4 Excitement	78	65	13	0.43	0.44
G6 Depression	79	62	17	0.30	0.40
N4 Passive/apathetic social withdrawal	90	81	9	0.35	0.36
G12 Lack of judgment and insight	85	76	9	0.26	0.27
P5 Grandiosity	81	71	10	0.21	0.22
G1 Somatic concern	65	55	10	0.96	0.93
G5 Mannerisms and posturing	71	81	10	0.18	0.24
P1 Delusions	91	56	35	0.17	0.24
N7 Stereotyped thinking	89	49	40	0.14	0.20
G2 Anxiety	89	77	12	0.14	0.20
G13 Disturbance of volition	91	57	34	0.13	0.17
G8 Uncooperativeness	78	71	7	0.11	0.12
G3 Guilt feelings	80	51	29	0.10	0.16

Disturbance of volition (G13)

Disturbance of volition was rated higher by Polish psychiatrists, probably at the expense of passiveness and apathy. Both items should be defined more precisely because passiveness and lack of interest may be interpreted as indecisiveness. The discrepancy between ratings may have been affected by comprehension

problems (the translated content was presented in subtitles that do not stress relevant differences).

Depression (G6), feelings of guilt (G3)

Depression (G6) and feelings of guilt (G3) received low ratings. In discussion, Polish trainees frequently emphasized that the discrepancy resulted from the

Table 5. Inter-rater agreement higher than rater-standard agreement

Symptoms	Percentage of inter-rater agreement A	Percentage of rater-standard agreement B	A–B	ICC
P1 Delusions	91	56	35	0.24
G13 Disturbance of volition	91	57	34	0.17
N7 Stereotyped thinking	89	49	40	0.20
G3 Feelings of guilt	80	51	29	0.16
G6 Depression	79	62	17	0.40
P7 Hostility	83	72	11	0.46
P5 Grandiosity	81	71	10	0.22
N4 Passive/apathetic social withdrawal	90	81	9	0.36
G12 Lack of judgment and insight	85	76	9	0.27

Table 6. Highest rater-standard disagreement (percentage of ratings)

Observer ratings exceeding standard ratings	+1	+2	> +2	Total
P1 Delusions	24	36	8	68
P4 Excitement	47	18	17	82
N7 Stereotyped thinking	30	42	9	81
G 13 Disturbance of volition	44	36	6	86
Standard ratings exceeding observer ratings	–1	–2	> –2	Total
G1 Somatic concern	11	16	29	56
G3 Feelings of guilt	23	36	11	70
G5 Mannerisms and posturing	34	27	11	72
G7 Depression	37	27	15	79
N4 Passive/apathetic social withdrawal	33	12	6	51

fact that authors attached greater importance to verbal declarations whereas the raters focused on the absence of depressive expression in the patient. These items should be supplemented with a comment stating that the rater must above all consider the patient's statements rather than his or her own impressions.

Somatic concern (G1)

The discrepancy in the evaluation of item G1 – concern for one's own physical health – resulted from the absurdity of the patient's delusions. This item should bear a comment stating that delusions of bodily change are regarded as hypochondriac, regardless of their likelihood.

As did other authors (Andreasen et al., 1982; Peratla and Cuesta, 1994; Müller et al., 2000), the

Polish study found that the level of professional expertise (seniority or degree of specialization) did not affect the accuracy of ratings.

Conclusions

The study demonstrated that Polish trainees' ratings of several items of the PANSS differ from the respective standard ratings. These items require additional comments addressing the differences in approach to assessment of the severity of pathologies. Comments should pertain to the following items:

- P1 delusions;
- G1 somatic concern;
- G3 feelings of guilt;
- G6 depression.

Differences in the rating of other symptoms may progressively decrease as a result of repeated training of raters. In my opinion, the remaining items do not need to be supplemented with additional comments in the PANSS manual.

Acknowledgements

I am indebted to Janssen-Cilag and especially to Dr Alina Kucma for providing recordings and organizing training sessions.

References

- Andreasen NC. Negative symptoms in schizophrenia: definition and reliability. *Arch. Gen. Psychiatry*, 1982; 39: 784–8.
- Kay SR, Opler LA, Lindenmayer JP. Reliability and validity of the Positive and Negative Syndrome Scale for schizophrenics. *Psychiatry Research* 1987; 23: 99–110.
- Kay SR. Positive-negative symptom assessment in schizophrenia: psychometric issues and scale comparison. *Psychiatr. Quarterly* 1990; 61: 163–78.
- Kay SR. *Positive and Negative Syndromes in Schizophrenia*. New York: Brunner & Mazel, 1991.
- Lambert T. Inter-rater reliability and the PANSS. Paper presented at the CINP Congress, Melbourne, Australia, 1996.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
- Lindström E, Wieselgren IM, Von Knorring L. Inter-rater reliability of the structured clinical interview for the Positive and Negative Syndrome Scale for schizophrenia. *Acta Psychiatr. Scand.* 1994; 89: 192–5.
- Müller MJ, Rossbach W, Dannigkeit P, Müller-Siecheneder F, Szegedi A, Wetzel H. Evaluation of standardized rater training for the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Research* 1998; 32: 151–60.
- Müller MJ, Rossbach W, Davids E, Wetzel H, Benkert O. Evaluation eines standardisierten trainings für die Positive and Negative Syndrome Scale (PANSS). *Nervenarzt* 2000; 71: 195–204.
- Overall JE, Gorham DR. The Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. *Psychopharmacol. Bull.* 1988; 24: 97–9.
- Peralta V, Cuesta MJ. Psychometric properties of the Positive and Negative Syndrome Scale (PANSS) in Schizophrenia. *Psychiatry Research* 1994; 53: 31–40.
- Von Knorring L, Lindström E. The Swedish version of the Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Construct validity and inter-rater reliability. *Acta Psychiatr. Scand.* 1992; 86: 463–8.

*Correspondence: Małgorzata Rzewuska, Assistant Professor, Institute of Psychiatry, Warsaw, Independent Unit of Pharmacotherapy, Sobieskiego 9, 02-957 Warsaw.
Telephone/fax (+48) 22 651-9312.
Email: rzewuska@ipin.edu.pl.*