

The Brief Psychiatric Rating Scale : Interrater Reliability

By: E. Hamdi, Y. Amin, R. Maftouz, O. Arafa and Y. Rakhawy

ABSTRACT

The application of psychiatric rating scales in follow-up and research should be preceded, among other prerequisites, by a study of interrater reliability. The Brief Psychiatric Rating Scale (B.P.R.S.) has been used extensively in research settings. This study attempted to establish interrater reliability for the B.P.R.S. in an Egyptian sample of psychiatric patients. Seventy one in-patients were selected at random and subjected to assessment of their symptoms by the B.P.R.S. They were assessed independently by four teams of raters, each team composed of two raters. Interrater reliability was studied for the 23 symptom areas of the B.P.R.S. using the Pearson Product Moment Correlation Coefficient. Among the four teams, adequate interrater reliability was established for 17 items (range 0.50 to 0.97). Moderate correlation (range 0.38 to 0.49) was found in one or two teams of raters in 6 items. A poor or negative correlation was found in one or two teams of raters in hallucinatory behavior, uncooperativeness, and mannerisms. Poor correlations were traced to minimal appearance of these symptoms leading to a statistical bias, as well as discrepancies in the knowledge of the raters about the patient's behavior outside the interview. In general reliability tends to be high whenever the item under con-

sideration relies upon the patient's verbal reports rather than upon the opinion of the rater.

INTRODUCTION

Psychiatric rating scales represent attempts to measure aspects of behaviour in standard and reliable ways. At the present state of knowledge, they are far from being precise tools. For the purpose of replicable research, the collection and comparison of data derived from large samples through computerized means, rating scales present a distinct advantage over global opinions and unsystematic assessment (Glueck and Stroebe, 1975).

The Brief Psychiatric Rating Scale (Overall and Gorham, 1962; Gorham and Overall, 1960, 1961) has been widely recognized and used both for the routine follow-up and research assessment of psychiatric patients, notably the assessment of change brought about by treatment (Overall, 1973), and in attempts to relate psychiatric symptoms to a number of independent variables (Ford, 1975).

According to Overall and Gorham (1962) "The primary purpose in developing the brief scale has been the development of a highly efficient, rapid evaluation procedure for use in assessing treatment change in psychiatric patients, while at the same time yielding a rather comprehensive description of major symptom characteristics". These advantages probably underlie the wide applicability of the B.P.R.S.

The original version developed by Overall and Gorham (1962) includes 16 symptom constructs separated by factor analyses. Several revisions have been made with the addition of more symptom areas. The version used in this work includes 23 operationally defined symptom areas (Bigelow and Murphy, 1979).

The task of assessment falls upon the clinician who has to assess the patient in each symptom area on a 7 - point scale.

Rating of symptom severity is both quantitative and qualitative, but the two dimensions are combined in a single score.

This study has been undertaken in order to implement the B.P.R.S. in routine patient assessment and follow-up at Dar El-Mokattam Hospital for Mental Health. It was considered necessary to conduct a preliminary study of interrater reliability so as to correct inconsistencies and differences of opinion in the rating of psychiatric symptoms.

METHOD

Pre-assessment Training:

Eight psychiatrists participated in this study (range of experience 3–11 years). Conjoint meetings were held in which the B.P.R.S. manual of operational definitions was reviewed, and discussed. This served the purpose of resolving differences in conceptualization, as well as delineating the boundaries of each symptom area. A process of conjoint assessment was then started, in which the 8 raters met and rated a number of patients independently after an interview conducted by a senior rater. The ratings of each item were then discussed, and disagreements resolved by consulting the scoring manual. At the end of the training meetings four teams were formed, each composed of a senior and junior psychiatrist.

Patients:

No criteria for selection were used, and patients were successively recruited from the in-patient population of Dar El-Mokattam. The distribution of clinical diagnoses of the patient sample is presented in table 1. Raters 1 & 2 assessed 50 patients, raters 3 & 4 assessed 42 patients, raters 5 & 6 assessed 26 patients, and raters 7 & 8 assessed 25 patients.

Assessment Interviews:

The assessment interview was conducted separately by each team. One of the two raters would conduct the interview, allowing for the second rater to clarify obscure points at the end. Both raters alternated in conducting the interviews. The interview time (approximately 20 minutes) was divided in the following manner: a standard introduction and explanation of the purpose, followed by 3 – 5 minutes of non-directive questioning about the patient's immediate concerns. Ten to fifteen minutes were dedicated to enquiry about the patient's complaints and circumstances of the illness. The last 5 to 10 minutes were focused upon direct questioning so as to fill the gaps in spontaneous symptom reporting by the patient. Scoring was done independently at the end of each interview. Raters were allowed during the initial interviews to exchange knowledge about their independent ratings without modifying them.

Statistical Analysis:

The Pearson Product Moment Correlation coefficient was calculated to estimate the degree of correlation between each pair of raters in each symptom area. The symptoms were then grouped into five domains covering anxiety, depression, paranoia, mania, and schizophrenia. The total scores in each domain were compared among each team of raters yielding a measure of interrater reliability in the assessment of the different syndromes covered by the B.P.R.S. Testing the significance of the estimated correlations was done by consulting the relevant tables (Guilford and Fruchter, 1978). Spearman Brown correction was employed on the mean correlations of hallucinatory behavior, mannerisms and posturing, and uncooperativeness to compensate for a low range of scores. The latter leads to a spurious lowering of the correlation coefficient (Hill, 1981).

TABLE I
Distribution of Clinical Diagnoses of Patient Sample
(DSM-III)

Diagnosis	N	%
1. Schizophrenic Disorders	25	35.5
2. Bipolar Disorder, Manic	8	
Mixed	5	21
Depressed	2	
3. Major Depression	6	8.4
4. Paranoid Disorders		
Acute	8	12.6
Paranoia	1	
5. Schizo-Affective Disorder	5	7
6. Atypical Psychosis	5	7
7. Anxiety Disorders	3	4.2
8. Schizophreniform Disorder	2	2.8
9. Schizotypal Personality disorder	1	1.4
Total	71	100

RESULTS

1 - Symptom Reliability:

Table II illustrates the estimates of interrater reliability among four couples of raters using the B.P.R.S. in an Egyptian sample of psychiatric inpatients. Raters 5 & 6 have the highest mean correlation (0.75) with a range between -0.08 to 0.94. All their correlations are statistically significant ($P < 0.005$) except helplessness/hopelessness ($p < 0.01$) and hallucinatory behavior (non-significant).

TABLE II

Interrater Reliability of B.P.R.S. Symptoms among Four Couples
of Raters in an Egyptian Sample of Psychiatric Patients

Item	Raters 1&2	Raters 3&4	Raters 5&6	Raters 7&8
1. Anxiety Statements	0.89	0.57	0.72	0.64
2. Tension (Behavior)	0.67	0.62	0.81	0.64
3. Depressive Mood	0.68	0.50	0.65	0.82
4. Helplessness/Hopelessness	0.61	0.41	0.49	0.87
5. Guilt Feelings.	0.77	0.71	0.89	0.88
6. Somatic Concern	0.77	0.55	0.82	0.61
7. Hostility	0.85	0.46	0.81	0.38
8. Suspiciousness	0.66	0.83	0.92	0.81
9. Uncooperativeness	0.56	0.50	0.69	-0.08
10. Distractability	0.78	0.46	0.76	0.75
11. Elated Mood	0.70	0.73	0.75	0.69
12. Motor Hyperactivity.	0.86	0.41	0.91	0.77
13. Disorientation	0.43	0.66	0.84	0.81
14. Disorganized Speech	0.76	0.47	0.80	0.93
15. Grandiose Statements	0.77	0.65	0.84	0.97
16. Unusual Thought Content	0.71	0.58	0.80	0.96
17. Hallucinatory Statements	0.88	0.77	0.94	0.92
18. Hallucinatory Behavior	0.83	0.20	-0.08	0.88
19. Emotional Withdrawal	0.75	0.61	0.60	0.65
20. Blunted Affect	0.74	0.70	0.69	0.77
21. Motor Retardation	0.85	0.56	0.81	0.67
22. Mannerisms & Posturing	0.50	0.54	0.80	-0.01
23. Loss of Functioning	0.66	0.50	0.75	0.82
Total Score	0.83	0.76	0.91	0.90

Raters 1 & 2 have the second highest mean correlation (0.73) with a range between 0.43 and 0.89. All their correlations are statistically significant beyond the 0.005 confidence limit.

Raters 7 & 8 have a mean correlation of 0.71 (range -0.1 to 0.97). With the exception of hostility ($P \leq 0.05$), uncooperativeness, and mannerisms (non-significant), all their correlations are significant beyond the 0.005 confidence limit.

Raters 3 & 4 have a mean correlation of 0.57 (range 0.20 to 0.83). With the exception of hallucinatory behavior which is non-significant, all their correlations are statistically significant beyond the 0.005 confidence limit.

Table III illustrates the overall mean correlation of the four couples of raters in each symptom area arranged in descending order. Hallucinatory statements reported by the patient have the highest interrater reliability (0.88) followed by the ratings of suspiciousness (0.81), guilt feelings (0.81), and grandiose statements (0.81). The least estimates of interrater reliability are those for uncooperativeness (0.58), helplessness/hopelessness (0.60), mannerisms and posturing (0.61), and hallucinatory behavior (0.63).

Table IV compares the estimates of interrater reliability obtained from the present study with those obtained by Overall and Gorham (1962) using the older version of the B.P.R.S. Although interrater reliabilities obtained by Overall and Gorham are on the whole higher than those obtained in our study, the difference is more than 0.1 in four symptoms, namely anxiety, hostility, mannerisms and posturing, and somatic concern. The overall mean correlation on tension in our study is higher than Overall and Gorham study by more than 0.1. In both studies, hallucinations have highest interrater reliability. Guilt feelings, grandiosity, and suspiciousness correlated beyond 0.8 in both studies.

TABLE III

Mean Interrater Reliability of B.P.R.S. Scores among four
Couples of Raters in an Egyptian Sample of Psychiatric Patients

Symptom	Mean Correlation
1. Hallucinatory Statements	0.88
2. Suspiciousness	0.81
3. Guilt Feelings	0.81
4. Grandiose Statements	0.81
5. Unusual Thought Content	0.76
6. Motor Hyperactivity	0.74
7. Disorganized Speech	0.74
8. Blunted Affect	0.73
9. Elated Mood	0.72
10. Motor Retardation	0.72
11. Anxiety Statements	0.71
12. Tension (Behavior)	0.69
13. Distractability	0.69
14. Disorientation	0.69
15. Somatic Concern	0.69
16. Loss of Functioning	0.68
17. Depressive Mood	0.66
18. Emotional Withdrawal	0.65
19. Hostility	0.63
20. Hallucinatory Behavior	0.63 *(0.46)
21. Mannerisms & Posturing	0.61 *(0.44)
22. Helplessness/Hopelessness	0.60
23. Uncooperativeness	0.58*(0.42)
Total Score	0.85

* Spearman-Brown correction done to compensate for minimum appearance (low range of scores).

TABLE IV
Estimates of Interrater Reliability Compared to
Overall and Gorham (1962)

Symptom	Overall and Gorham	Hamdi et al.
1. Hallucinations	0.87	0.88
2. Guilt Feelings	0.87	0.81
3. Anxiety	0.86	0.71*
4. Hostility	0.86	0.63*
5. Grandiosity	0.84	0.81
6. Suspiciousness	0.84	0.81
7. Mannerisms and Posturing	0.84	0.61*
8. Unusual Thought Content	0.83	0.76
9. Somatic Concern	0.81	0.69*
10. Conceptual disorganization	0.80	0.74
11. Depressive Mood	0.76	0.66
12. Motor Retardation	0.72	0.72
13. Uncooperativeness	0.68	0.58
14. Blunted Affect	0.67	0.73
15. Emotional Withdrawal	0.62	0.65
16. Tension	0.56	0.69*
Mean Correlation	0.78	0.72

* Reliability differs by more than 0.1 between the two studies.

TABLE V
Estimates of Interrater Reliability in the Assessment of
Psychiatric syndromes Covered by the B.P.R.S.

Syndrome	Raters 1&2	Raters 3&4	Raters 5&6	Raters 7&8
1. Anxiety:				
Anxiety Statements	0.85	0.67	0.79	0.60
Tension	Overall Correlation = 0.73			
2. Depression:				
Depressive mood	0.82	0.73	0.92	0.89
Helplessness				
Guilt Feelings				
Somatic Concern				
Motor Retardation				
	Overall Correlation = 0.84			
3. Mania:				
Distractability				
Elated Mood	0.88	0.60	0.92	0.83
Motor Hyperactivity				
Grandiose Statements				
	Overall Correlation = 0.81			
4. Paranoia:				
Hostility				
Suspiciousness	0.81	0.69	0.80	0.70
Uncooperativeness				
	Overall Correlation = 0.75			
5. Schizophrenia:				
Disorganized Speech				
Unusual Thought Content				
Hallucinatory Statements	0.75	0.77	0.88	0.90
Hallucinatory Behavior				
Emotional Withdrawal				
Mannerisms and Posturing				
	Overall correlation = 0.83			

II – Syndrome Reliability:

Table V presents the estimates of interrater reliability in the assessment of five psychiatric syndromes represented in the B.P.R.S. symptoms. The highest overall correlations among the four couples of raters is that on symptoms related to depression (0.84), while the lowest correlation is that on symptoms related to anxiety (0.73). The interrater reliability of symptoms related to schizophrenia, mania, and paranoia are 0.83, 0.81, and 0.75 respectively.

DISCUSSION

Our study demonstrates an adequate overall interrater reliability among most individual items of the B.P.R.S. Estimates of reliability are similar, although slightly lower than those reported by Overall and Gorham (1962). Our estimates are also similar to those reported by Peck and Dean (1983) who gave an interrater reliability ranging from 0.56 to 0.87. Our mean correlations range between 0.58 and 0.88.

The lower estimates of reliability in some symptoms; namely uncooperativeness (0.58), helplessness/hopelessness (0.60), mannerisms (0.61), and hallucinatory behavior (0.63) is probably due to their infrequent appearance in our sample of patients. According to Roscoe (1969), the Pearson product moment correlation coefficient is spuriously lowered by a low range of scores in the sample studied. Actually in our study these items showed little disagreement among the raters, but most patients were rated as 0 (absent). The same explanation may underlie the discrepancy in the ratings of mannerisms between our sample (0.61) and the study of Overall and Gorham (1962). The latter sample was composed exclusively of schizophrenics who are expected to exhibit mannerisms more frequently than our heterogeneous sample.

The symptoms covered by the B.P.R.S. are divided into those that rely in their scoring mainly upon the patient's verbal reporting, and those symptoms that depend upon the evaluation and opinion of the interviewer. In our study, the former type of symptom was more reliable in its assessment than the latter. Thus, hallucinatory and grandiose statements, suspiciousness and guilt feelings, which all depend upon the patient's verbal report, were highest in terms of interrater reliability. On the other hand, items that require considerable inference and judgement on the part of the clinician had a lower reliability e.g. depressive mood, emotional withdrawal, hostility, and loss of functioning. The same phenomenon is apparent in the Overall and Gorham study where emotional withdrawal, tension, and blunted affect have the lowest reliability.

Although the estimates of reliability in our study are satisfactory, we should comment that at least part of the discrepancies in the raters' judgments came from a similar discrepancy in the amount of knowledge the raters had about the patient outside the interview. The junior rater usually spent more time with the patient in hospital activities, and usually had impressions about symptoms that were sometimes different from those of the senior rater.

Lastly, our study demonstrated that there is no great difference in the assessment of symptoms related to a particular diagnostic entity. The difference in overall correlation in assessing symptoms of depression (0.84) and anxiety (0.73) is 0.11 (table V). It is reasonable to assume that the reliability of using the B.P.R.S. in reaching a particular diagnosis included in its symptoms is also high.

The training and interview guidelines recommended by Overall and Gorham (1962) should be followed in order to enhance interrater agreement in using the B.P.R.S. We would like to add that a proper exchange of information about the patient prior to the rating interview is also important.

REFERENCES

- Ford, H. (1975) Involitional melancholia. In Freedman, A.M., Kaplan, H.I. and Sadock, B.J. (eds.), **Comprehensive Textbook of Psychiatry**. 2nd ed., Vol. 1. Baltimore: The Williams and Wilkins Co.
- Glueck, B.C. and Stroebe, C.F. (1975) Computers and clinical psychiatry. In Freedman, A.M., Kaplan, H.I. and Sadock, B.J. (eds.), **Comprehensive Textbook of Psychiatry**. 2nd ed., Vol. 1. Baltimore: The Williams and Wilkins Co.
- Gorham, D.R. and Overall, J.E. (1960) Drug action profiles based upon an abbreviated psychiatric rating scale. **J. Nerv. Ment. Dis.**, **132**: 528—535.
- (1961) Dimensions of change in psychiatric symptomatology. **Dis. Nerv. Syst.**, **22**: 576—580.
- Guilford, J.P. and Fruchter, B. (1978) **Fundamental Statistics in Psychology and Education**. (6th ed.) Toronto: Mc Graw Hill.
- Hill, A.B. (1981) **A Short Textbook of Medical Statistics**. Kent: Hodder and Stoughton Ltd.
- Overall, J.E. (1973) The brief psychiatric rating scale in psychopharmacologic research. In Pichot, P. and Oliver Martin, R. (eds.), **Psychological Measurements in Pharmacopsychiatry**. Basel: S.Kruger.
- Overall, J.E. and Gorham, D.R. (1962) The brief psychiatric rating scale. **Psychological Reports**, **10**: 799—812.
- Peck, D.F. and Dean C. (1983) Measurement in psychiatry. In Kendell, R.E. and Zealley A.K. (eds.), **Companion to Psychiatric Studies**. 3rd ed. London: Churchill Livingstone.
- Roscoe, J.T. (1969) **Fundamental Research Statistics**. New York: Holt, Rinehart and Winston.

AUTHORS

E. Hamdi M.B, Ch. B., M. Sc., M.D. (Psychiat.), Lecturer, Department of Psychiatry, Cairo University.

Y. Amin, M.B. Ch. B., M. Sc., M.D. (Psychiat.), Lecturer, Department of Psychiatry and Neurology, Suez Canal University.

K. Mahfouz, M.B. Ch.B., D.M., M.Sc., M.D. (Psychiat.), Director, Dar El-Mokattm Hospital for Mental Health.

O. Arafa, M.B. Ch.B., Resident in psychiatry, Dar El-Mokattam Hospital for Mental Health.

Y. Rakhawy, M.B. Ch.B., D.M., D.P.M. & N, M.D. (Psychiat.), M.R.C. Psych., Professor of Psychiatry, Cairo University.

الموجز

مقياس التقييم الطبني المختصر :

دراسة لمعدل الثبات بين المقيمين

عند تطبيق مقياس التقييم الطبني في البحث والمتابعة يجب أن تكون مسبوقة بدراسة للثبات بين المقيمين . وقد تم تطبيق المقياس الطبني المختصر في الأبحاث على مدى واسع ، وتحاول هذه الدراسة تحديد درجة الثبات بين المقيمين لهذا المقياس المطبق على عينة من المرضى النفسيين المصريين . وقد تم اختيار ٧١ مريضاً مقيمين بالمستشفى عشوائياً وقيمت أعراضهم باستخدام هذا المقياس بواسطة أربعة أزواج من المقيمين . ودرس معدل الثبات بين المقيمين لثلاث وعشرين عرضاً في المقياس الطبني المختصر باستخدام مكافئ معامل ارتباط ناتج بيرسون . وظهر معامل ارتباط كاف للثبات بين الفرق الأربعة في ١٧ نقطة (مدى ٠.٥٠ الى ٠.٩٧) . ومعامل ارتباط متوسط بين فريق أو اثنين من المقيمين في ست نقاط (مدى ٠.٣٨ الى ٠.٤٩) . وقد ظهر معامل ارتباط ضعيف أو سلبي بين فريق أو اثنين من المقيمين في نقاط السلوك الهلوسي وعدم التعاون والنمطية . وفسر ذلك بقلّة ظهور هذه الأعراض مما أدى الى تحيز احصائي وكذلك

الى اختلاف درجة معرفة المقيمين بسلوك المرضى خارج المقابلة . وأخيرا وجد أن معدلات الثبات تميل الى الارتفاع بوجه عام حينما يعتمد التقييم على ما يقرره المريض لفظيا أكثر مما تكون حينما يعتمد التقييم على رأى المقيم .

ABSTRAIT

Brief Psychiatric Rating Scale: Sa Fiabilité dans L'Evaluation Clinique

L'emploi des échelles dans l'évaluation clinique des symptômes dans les cas de maladies à suivre et dans les recherches doit être nécessairement précédé par une étude de leur fiabilité. En effet, la "Brief Rating Scale" a été longuement étudiée dans des contextes de recherche. L'étude présente essaie d'établir la fiabilité de la B.P.R.S. dans l'ealuation des symptômes dans un échantillon de malades Egyptiens.. Soixante onze malades hospitalisés ont été choisis par hasard et sujets à l'évaluation. Ces malades ont été examinés et évalués par quatre groupes de cliniciens, chaque groupe étant composé de deux cliniciens. La fiabilité dans chaque paire de cliniciens a été étudiée dans 23 zones de symptômes de la B.P.R.S. Le coefficient de 'Pearson Product moment correlation' a été utilisé dans ce but. Parmi les quatre groupes, une fiabilité suffisante a été établi pour chaque paire pour 17 symptômes (de 0.50 à 0.97). Une corrélation modérée a été trouvée dans un de deux groupes de cliniciens pour 6 symptômes. Une corrélation pauvre et négative a été trouvée dans un des deux groupes de cliniciens pour le comportement hallucinatoire, le peu de coopération et le maniérisme. Les corrélations pauvres ont été attribuées à l'apparition minimale de ces symptômes amenant à une fausse contre faveur statistique, et à la discordance dans le savoir des cliniciens concernant le comportement des malades hors de l'entrevue. En général, la fiabilité tend a être haute quand le point considéré dépend plus sur les paroles du patient que sur l'opinion du clinicien.

مصحة النيل بالمعادي

للأمراض النفسية والعصبية

٢٤٠ طريق مصر حلوان بالمعادي

ت : ٥٠١٦٧١ - ٥٠١٤٣٣ - ٥٠١٥٤٢

تأسست عام ١٩٥٢

أقسام لعلاج الحالات النفسية . والحالات العصبية
أقسام خاصة لرعاية وعلاج المسنين . قسم للأطفال

قسم خاص لعلاج حالات الإدمان

تقسم للمرضى جميع أنواع العلاجات المختلفة

العلاج بالعمل

ترفيه - رحلات - حضانات

الإدارة والإشراف الطبي

مشارطة الطب النفسي والأعصاب
عضو كلية الأطباء النفسيين الملكية

يتعاون في الإشراف على المرضى

تخبة كبيرة من أساتذة الطب النفسي

والأخصائيين بجمهورية مصر العربية