The Brief Psychiatric Rating Scale for Children: A Reliability Study

John Gale, Betty Pfefferbaum, Melanie A. Suhr, and John E. Overall

Department of Psychiatry
University of Texas Health Sciences Center at Houston

Forty-eight children and adolescents with widely varied presenting problems were interviewed by paired clinician-raters to evaluate the reliability of the Brief Psychiatric Rating Scale for Children (Overall & Pfefferbaum, 1982), a 21-item rating scale that provides a descriptive profile of child and adolescent psychopathology. Eighteen of the 21 scales appeared adequately reliable when examined individually. All seven composite factor scores proved acceptably reliable, suggesting the advantage of using those composite scores for the evaluation of treatment effects in the child and adolescent patient populations.

Key words: Brief Psychiatric Rating Scale for Children (BPRS-C), rating scale for children, reliability

Recently, the Brief Psychiatric Rating Scale for Children (BPRS-C) was introduced into the literature with the aim of providing an efficient, broader, more flexible instrument for clinical research in child and adolescent patient populations (Overall & Pfefferbaum, 1982). It is intended for comprehensive, yet parsimonious, assessment of major symptom characteristics of young patients. Based on factor analysis of an extended set of symptom-descriptive items (Pfefferbaum & Overall, 1983), successive sets of three BPRS-C scales provide quantitative representation of seven major dimensions of psychopathology, separating some 18 DSM-III (American Psychiatric Association, 1980) child and adolescent diagnostic groups. The BPRS-C is similar to the Brief Psychiatric Rating Scale (BPRS) for adults (Overall & Gorham, 1962) in providing a format for rating the presence and severity of 21 symptoms on individual 7-point scale items ranging from not present to extremely severe.

The aim of the BPRS-C, as embodied in the rating scale format provided in the Appendix to this article, is somewhat different from that of existing instruments intended for use in younger patient populations. Specifically, the BPRS-C is not designed for intensive assessment of a narrow aspect of psychopathology or a particular syndrome, nor is it primarily intended as a diagnostic tool. Rather, it was developed to yield a concise descriptive profile applicable to a broad range of child and adolescent psychopathology (Overall & Pfefferbaum, 1982; Pfefferbaum & Overall, 1982). It was developed to include representative symptoms and behaviors for the full range of DSM-III child and adolescent diagnoses, so that few young patients should defy quantitative description of the major aspects of their psychopathology by this brief rating instrument. It is considered an appropriate tool for (a) investigations of differences in clinical manifestations related to demography, epidemiology, or treatment modality; (b) descriptive rather than diagnostic classification; and (c) longitudinal following of treatment response in both naturalistic and experimental research paradigms. These are all ways in which the adult BPRS has been used in clinical research (Guy, 1976; Hedlund & Vieweg, 1980), and they appear equally relevant for the BPRS-C, which has had a similar research history.

The purpose of this article is to report a study of interrater reliability of the BPRS-C scale items and factor scores. The interrater method, which was used to study BPRS-C reliability, requires two clinician-raters to rate each subject independently. The product-moment correlation between paired independent ratings of the same subjects provides a numerical index of the consistency with which the clinical phenomena are observed and evaluated. The product-moment coefficient was selected for
use in assessing consistency between raters because, under tenable assumptions, such coefficients most directly represent the proportion of total variation or differences between patients that is agreed on by the raters. In the case of clinical ratings, lack of consistency between independent ratings of the same subjects is considered to represent measurement error, whatever the cause.

**Method**

**Subjects**

Subjects were 48 children. Twenty-eight children were seen at the Houston Child Guidance Center, which is a public multidisciplinary outpatient clinic serving Houston and surrounding areas. These subjects were taken in order, as work load permitted, from a self- or family-referred waiting list. Participation in the study was incorporated into the initial psychiatric screening evaluations which routinely require patient and family interviews as part of the intake process of the clinic. The remaining 20 children were seen at the Child and Adolescent Psychiatric Inpatient Unit of Hermann Hospital, which is the primary teaching hospital of the University of Texas Medical School at Houston. These subjects were consecutive admissions to the staff teaching service. No attempt was made to select or exclude certain types of patients from the study at either institution.

The subjects (33 boys, 15 girls) ranged in age from 5 to 18 years. On the whole, a good distribution of ages was present at both interview sites, although more of the adolescents were seen at the inpatient unit of Hermann Hospital. Boys outnumbered girls approximately 2:1 in both clinical settings. The subjects represented a wide range in socioeconomic status (SES), which tended to be correlated with site of interview. The Houston Child Guidance Center serves a largely lower middle and lower SES population, with a sliding pay scale for services. The child and adolescent unit at Hermann Hospital functions as a private inpatient service with a largely middle-class population and fees most often covered by third party.

By including both inpatient and outpatient subjects in our study, we achieved a subject sample spanning a broad range of presenting problems: 20 children (42%) with primary complaints of a depressed, fearful, and/or anxious nature; 19 children (40%) with primary problems of an antisocial, aggressive, and/or oppositional nature; 7 children (14%) with primary complaints of a psychotic or severely developmentally disturbed nature; and 2 children (4%) for reasons associated with environmental pressures, such as custody dispute. No attempt was made to select unambiguous diagnostic prototypes as subjects in the study. The subjects came to the interview early in their hospital or outpatient evaluation, without prior screening.

**Procedure**

Three of the authors of this report (Gale, Pfefferbaum, & Suhr) worked in pairs to interview and rate psychopathology for a total of 48 child subjects. The subjects were apportioned equally among the three possible pairings of the raters so that each clinician rated 32 of the 48 subjects, 16 with each of the other two raters. This apportionment of subjects was accompanied by systematically rotating the various rater pairings so that each rater was paired with each other rater an equal number of times throughout the course of the study. Two of the three clinician-raters were psychiatrists holding faculty positions in the Division of Child Psychiatry. The other was a fourth-year fellow in child psychiatry. All are considered to have specialization appropriate for reliable use of the BPRS-C. Rater pairs rotated in the interview assignment for consecutive admissions. Inspection of the age, sex and presenting symptoms seen by each clinician-rater pair revealed that for the most part each pair interviewed similar sample subgroups. The only exception was the pair of Pfefferbaum and Suhr, who saw a moderately high number of boys (8) in the 15- to 18-year age category.

In each setting, the typical interview session lasted approximately 50 min and was divided into an interview first with the accompanying adults and then with the child. The format was used whenever possible, although clinical circumstances necessitated occasional deviation. Both clinician-raters of a pair were present for the entire interview. The role of primary interviewer was systematically alternated within each pair, with the other clinician-rater observing passively until the end of the session when he or she was permitted to engage the subject in further brief exchange for the purpose of clarifying any issues as needed.

Immediately after a session, the two clinician-raters independently completed the BPRS-C ratings (see Appendix). After that, they sometimes discussed the case in general terms, usually with the clinical aim of devising the best disposition for a child seen at the Houston Child Guidance Center; however, discussion of specifics relating directly to the BPRS-C were avoided throughout the study. Ratings were not discussed nor compared at any time during the data acquisition phase of the study.

**Analysis of Data**

A simple Pearson product–moment correlation coefficient was calculated between paired ratings across the 48 subjects for each of the 21 scale items.
The product-moment correlation between two independent measurements of the same thing is an estimate of the proportion of total variation in either one of the sets of scores that is not due to error. That is, the coefficients obtained by correlating ratings by two independent clinician-raters pertain to the reliability of ratings by a single rater.

The Spearman-Brown prophecy formula can be used to provide an estimate of the reliability of measurements that are the sums or averages of ratings by two independent raters. This is analogous to estimating the reliability of a test that is twice as long as a shorter test for which reliability is known.

Factor scores for the BPRS-C consist of the sums of ratings on subsets of three related scale items. Coefficient-alpha internal consistency estimates of factor score reliability were calculated from intercorrelations among scores on the three scales in each factor domain. Estimates of reliability for factor scores averaged across two independent raters were then extrapolated using the Spearman-Brown prophecy formula, as was done for the individual scale items.

Results

Reliability coefficients estimated by the interrater method for the individual items of the BPRS-C are presented in the column headed \( r_2 \) in Table 1. These are estimates of the extent to which ratings provided by a single rater reflect true differences among patients. The entries in the column labeled \( r_{s2} \) are Spearman-Brown estimates of the reliability of scores resulting from summing or averaging ratings provided by two independent observers.

The individual scale items of the BPRS-C are grouped in Table 1 according to the primary factor to which each relates. The factor constructs are identified by name immediately to the right of the individual item reliabilities. The alpha reliability coefficients for the factor scores are provided in the column of Table 1 labeled \( r_{cc} \). Again, these are estimates of the reliabilities that one would expect for factor scores obtained by summing ratings provided by a single clinician-rater. The Spearman-Brown estimates of reliabilities of factor scores obtained by summing or averaging ratings provided by two independent raters are shown in the righthand column of the table.

Discussion

Three somewhat different concepts of reliability are represented in the results presented here. Interrater reliability is concerned with agreement between raters. It represents a correlation/regression definition of reliability. The interrater reliability coefficient is understood to be the proportion of total variance that can be accounted for by regression of observed scores on true values for the individuals being rated. As noted, the correlation coefficient calculated between ratings from two raters is an estimate of the reliability of ratings by a single rater.

The Spearman-Brown formula which was used to estimate the reliability achieved by summing or averaging ratings made by two independent raters is based on a variance components model. Accord-

<table>
<thead>
<tr>
<th>Symptom Scale</th>
<th>( r_{11} )</th>
<th>( r_{22} )</th>
<th>Composite Factor</th>
<th>( r_{cc} )</th>
<th>( r_{2cc} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncooperativeness</td>
<td>.465</td>
<td>.642</td>
<td>Behavior Problems</td>
<td>.773</td>
<td>.872</td>
</tr>
<tr>
<td>Hostility</td>
<td>.458</td>
<td>.627</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manipulativeness</td>
<td>.756</td>
<td>.861</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive Mood</td>
<td>.736</td>
<td>.848</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feelings of Inferiority</td>
<td>.472</td>
<td>.641</td>
<td>Depression</td>
<td>.818</td>
<td>.900</td>
</tr>
<tr>
<td>Suicidal Ideation</td>
<td>.850</td>
<td>.919</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peculiar Fantasies</td>
<td>.615</td>
<td>.762</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delusions</td>
<td>.882</td>
<td>.937</td>
<td>Thinking Disturbance</td>
<td>.871</td>
<td>.931</td>
</tr>
<tr>
<td>Hallucinations</td>
<td>.709</td>
<td>.830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.689</td>
<td>.816</td>
<td>Motor Agitation</td>
<td>.845</td>
<td>.916</td>
</tr>
<tr>
<td>Distactibility</td>
<td>.714</td>
<td>.833</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech or Voice Pressure</td>
<td>.673</td>
<td>.805</td>
<td>Withdrawal</td>
<td>.846</td>
<td>.917</td>
</tr>
<tr>
<td>Underproductive Speech</td>
<td>.720</td>
<td>.837</td>
<td>Anxiety</td>
<td>.689</td>
<td>.816</td>
</tr>
<tr>
<td>Emotional Withdrawal</td>
<td>.634</td>
<td>.776</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blunted Affect</td>
<td>.616</td>
<td>.762</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tension</td>
<td>.642</td>
<td>.782</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>.544</td>
<td>.705</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep Difficulties</td>
<td>.668</td>
<td>.801</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorientation</td>
<td>.815</td>
<td>.898</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech Deviance</td>
<td>.890</td>
<td>.942</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereotypy</td>
<td>.748</td>
<td>.856</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Estimated Reliabilities for BPRS-C Symptom Scales and Composite Scores

Journal of Clinical Child Psychology, Volume 15, Winter, 1986 • 343
ing to that model, reliability is the ratio of true variance to total variance for a set of scores. The total variance of a linear combination of observed scores is a quadratic form which reduces to a weighted sum of elements in the matrix of intercorrelations among the observed scores and reliabilities of the observed scores. In the present case, reliabilities of the observed scores were estimated from the correlation between raters. Thus, the ratio of true variance to total variance for scores that are the sum of ratings by two independent raters is defined entirely by the interrater correlation.

The third concept of reliability that was invoked in this investigation derives from internal consistency among items contributing to a factor score. Coefficient alpha is frequently described as the mean of all possible split-half coefficients resulting from different ways of partitioning a set of items. This type of reliability reflects the consistency in scores on items that are supposed to measure the same thing. It is calculated from intercorrelations among scale items, not from correlations between raters. In this case, item intercorrelations were calculated across 32 x 3 sets of ratings because each of the three raters rated 32 patients. Extrapolation to reliability of factor scores averaged across raters was achieved by application of the Spearman-Brown formula to the coefficient-alpha results.

The results presented in Table 1 confirm an acceptable level of reliability for the seven factor scores, with the possible exception of scores on the Anxiety factor. If one were to go to the trouble and expense of having two experienced clinical observers to rate each patient, the factor scores obtained from their combined ratings would seem reliable enough for any purpose. The factor scores also tend to have better statistical distributions than do scores on individual scale items, although one cannot expect most patients to evidence pathology in all areas. Thus, substantial numbers of zero scores are to be expected, even on the factors.

Scores on scale items tend to be less reliable than factor scores, but they may be useful for specific purposes. For example, it may become important to know whether a patient had suicidal ideation, rather than simply a composite Depression factor score. Scale items are important, of course, for detailed profile analysis. For investigations where scale items are of primary interest, the use of more than a single rater might be considered as a way of enhancing reliability of the profile elements.

Summary

An investigation of the reliability of clinical assessments of manifest psychopathology in child and adolescent patients was conducted by three experienced clinicians using the BPRS-C. The clinician-raters worked in pairs in a balanced design in which each rater participated with each other rater in the interview and rating of 16 patients, yielding a total of 48 paired patient evaluations. Interrater reliability coefficients for the individual symptom ratings calculated across the 48 subjects ranged from .458 (hostility) to .882 (delusions) and .890 (speech deviance). It is common practice, where high reliability is required, to average the ratings by two independent raters. Spearman-Brown estimates of the reliability of symptom ratings obtained by averaging ratings from two independent raters ranged from .627 (hostility) to .937 (delusions) and .942 (speech deviance).

Factor scores assess more global aspects of behavior and generally have the advantage of increased reliability. The BPRS-C provides scores for seven relatively independent factors, each defined as the sum of ratings on three related items. For factor scores calculated from ratings by a single clinician rater, the internal consistency reliabilities ranged from .689 (Anxiety) to .871 (Thinking Disturbance) and .906 (Organicity). Factor scores calculated by averaging ratings from two independent observers ranged from .816 (Anxiety) to .931 (Thinking Disturbance) and .951 (Organicity). We conclude that these reliability estimates are generally quite adequate to support the clinical and research utility of the BPRS-C. As noted, some further work in defining criteria for rating anxiety manifestations may be worthwhile.

References


Received: 7/22/85
Revised: 1/27/86
1. Uncooperativeness - negative, uncooperative, resistant, difficult to manage.

2. Hostility - angry or suspicious affect, belligerence, accusations and verbal condemnations of others.

3. Manipulativeness - lying, cheating, exploitative of others.

4. Depressive Mood - sad, tearful, depressive demeanor.

5. Feelings of Inferiority - lacking self-confidence, self-deprecatory, feeling of personal inadequacy.

6. Suicidal Ideation - thoughts, threats, or attempts of suicide.

7. Peculiar Fantasies - recurrent, odd, unusual, or autistic ideations.

8. Delusions - ideas of reference, persecutory or grandiose delusions.

9. Hallucinations - visual, auditory, or other hallucinatory experiences or perceptions.

10. Hyperactivity - excessive energy expenditure, frequent changes in posture, perpetual motion.

11. Distractibility - poor concentration, shortened attention span, reactivity to peripheral stimuli.

12. Speech or Voice Pressure - loud, excessive, or pressured speech.

13. Underproductive Speech - minimal, sparse inhibited verbal response pattern, or weak low voice.

14. Emotional Withdrawal - unsympathetic relations to examiner, lack of peer interaction, hypoactivity.

15. Blunted Affect - deficient emotional expression, blankness, flatness of affect.

16. Tension - nervousness, fidgetiness, nervous movements of hands or feet.

17. Anxiety - clinging behavior, separation anxiety, preoccupation with anxiety topics, fears or phobias.

18. Sleep Difficulties - inability to fall asleep, intermittent awakening, shortened sleep time.

19. Disorientation - confusion over persons, places or things.

20. Speech Deviancy - inferior level of speech development, underdeveloped vocabulary, mispronunciations.

21. Stereotypy - rhythmic, repetitive, manneristic movements or posture.

1981 Overall and Pfefferbaum
