REHAB: A New Assessment Instrument for Chronic Psychiatric Patients

by Roger Baker and John N. Hall

Abstract

The renewed interest in the care and treatment of chronic psychiatric patients has led to an awareness of the need for improved assessment methods for this group. This article describes the development of REHAB, a behavior rating scale for use with people with chronic psychiatric disability, which has been carefully designed with respect to content, format, and ease of use. The article also draws together the various factor-analytic, reliability, validity, and sensitivity-to-change studies that provide the basic psychometric evidence for the scale's value. REHAB is presented as an integrated package of materials with special attention paid to training of raters and simplicity of interpretation. It has been used for a variety of different assessment purposes, in both research and clinical settings.

Over the last two decades there have been dramatic changes in the provision of services for chronic psychiatric patients both in the United States and Europe. The shift toward community care and the search for alternatives to mental hospital treatment (Stein and Test 1978; Flynn and Nitsch 1980) have, along with the undoubted benefits to patients, brought a new range of problems into focus. These include the differing needs of the new chronic population (Mann and Cree 1976), the needs of the most severely disabled patients (House of Commons Select Committee 1985), and the harmful effects of overzealous deinstitutionalization (Bachrach 1980). On both sides of the Atlantic, the continuing interest in the chronic psychiatrically disabled is indicated by the spate of new textbooks on the topic (Barofsky and Budson 1983; Lamb 1983; Watts and Bennett 1983; Caton 1984; Shepherd 1984).

A number of treatment approaches are now available for chronic or significantly disabled psychiatric patients. The substantial study of Paul and Lentz (1977) provided detailed information about both "social learning" and "milieu" approaches to treatment, while there continues to be interest in pharmacological approaches, such as minimal maintenance medication (Carpenter and Heinrichs 1983; Kane 1983). All of these approaches can have complex effects on individual patients, and hence require careful evaluation (Docherty 1984). Similarly, deinstitutionalization as a process has major social and psychological consequences, but there has been relatively little careful evaluation of the outcome of this process (Test and Stein 1978; Braun et al. 1981). A Maryland study of new chronic patients suggests there is "a paucity of [assessment of new long-stay] patients in the United States" (Platman 1983, p. 606).

As far as assessment of individual patients is concerned, there are two main methods in common clinical use. First are the interview-based rating scales completed on the basis of a more or less structured interview, such as the widely used Brief Psychiatric Rating Scale (BPRS) (Overall and Gorham 1962), although some other brief scales have been developed for use with chronic patients (Krawiecka et al. 1977;
Heinrichs et al. 1984).

Second are the behavioral rating scales, completed on the basis of a period of observation by direct-care staff such as nurses, of which the Nurses’ Observation Scale for Inpatient Evaluation (NOSIE) (Honigfeld and Klett 1965) is a widely used example. Some scales may combine both interview-based and observation-based approaches such as the modification of the BPRS to a 27-item version used by nurses (Kleinman et al. 1982). The advantages of observation-based scales are that they are based on the judgments of the care staff most in contact with the patient, are relatively quick to complete, and measure the sort of spontaneously occurring behavior that is relevant to everyday living. Reviews of different types of assessment for the severely psychiatrically disabled can be found in Weissman (1975), Hall (1979), Wallace (1981), Weissman et al. (1981), Anthony and Farkas (1982), and Beels et al. (1984). A recent article by Wallace (1986) in this journal reviews a number of functional assessment scales, including REHAB, the instrument described in this article.

There is no one perfect method of assessment for chronic patients, and to the extent that different types of assessment examine different facets of a patient’s behavior, more than one type of assessment may be required (Snith 1981). Certainly, the relationship between rating-scale data and data obtained from behavioral time-sampling techniques needs to be examined closely (Woods et al. 1984). A review of 17 behavior rating scales published in the United States, plus 12 scales published elsewhere, indicated that only four of these scales met five minimal criteria relating to their construction and evaluation (Hall 1980).

Despite the widespread use of observation-based scales and their obviously high face validity, the quality of the available instruments could be improved. There is scope for an observation-based scale that is technically adequate and at the same time could be used in the relatively poorly staffed environment in which many chronic patients live. This article reports the development of such a scale.

REHAB (Rehabilitation Evaluation Hall and Baker) (Baker and Hall 1984) is a 23-item behavior rating scale, containing a 7-item Deviant Behavior (DB) subscale and a 16-item General Behavior (GB) subscale, which is itself divided into five factors. The scale was designed for use with people with a chronic or disabling psychiatric handicap, who at the time of assessment are living in, or attending, a residential or day-care institutional setting. The term “patient” is used here to describe the people being rated, although the scale is applicable to residents in nonhospital settings.

In addition to the rating form, the scale is presented as an integrated package of materials including a detailed “User’s Manual” that explains how to set up ratings, train raters, and score and interpret the scale; it also provides technical data on score distributions, factorial structure, and normative data. In addition, the package includes separate “Rater’s Guides,” which explain the scale to the direct-care staff who make the ratings. There are “Score Sheets” to store data as well as individual and group data “Presentation Sheets” to display the results for the benefit of a decision-making team, or for the patients’ case notes.

**Development of the Scale**

**Objectives.** From 1970 to 1977 a major research project into the effectiveness of operant conditioning techniques with chronic psychiatric patients was mounted by the Department of Psychiatry and Psychology of the University of Leeds, England (Baker et al. 1974, 1977; Hall et al. 1977). Early in that research the limitations of existing behavioral rating scales became apparent. Our objectives were to develop a new scale that could:

- Be used in a range of settings, so that differing units within a treatment system, or different systems, could be compared.
- Generate a measure of general disability for overall ranking or grading of patients.
- Be capable of identifying patients presently living in an institution who had potential for living in the community.
- Identify general targets for treatment, so the most important or frequent problems of an individual or group of patients could be addressed.
- Be sensitive to change. Conventional criteria of reliability and validity do not incorporate sensitivity to therapeutic change, so this aspect of a scale requires separate examination.
- Be capable of being used repeatedly over a considerable period of time, with possibly several different raters, to evaluate long-term effects of psychological or milieu treatments or of drugs.

**Content.** The content of many rating scales is determined primarily by subjective means—usually professional opinion or reference to the items chosen in previous
scales. Also the content that is initially chosen often remains unchanged despite indications that some items should be discarded on the grounds of poor reliability, poor discrimination, or low factorial weighting. Substantial variations in content between ward rating scales ostensibly covering the same area are thus found (Hall 1977). We attempted to develop a more objective methodology for item selection, applying it to the initial item selection for REHAB. The method is described in detail by Baker (in preparation), but in brief it involved the following steps:

1. Choosing "source studies" that would provide a core of information on which to base item selection. Nine studies that had been conducted with long- and short-stay psychiatric or schizophrenic patients were chosen. Together, they (a) provided information relevant to all the various objectives set for the scale; (b) used three different methods for analyzing behavior (statistical, frequency count, and opinion survey); (c) referred to work on hospital- and community-based populations; and (d) referred to work carried out in the United States and the United Kingdom. Studies had to be well-constructed or, in the case of surveys, large.

The following source studies were selected: Two U.S. (Linn 1970; Strauss and Carpenter 1974) and two U.K. (Morgan and Cheadle 1974; Hall et al. 1977) studies provided information on the statistical relationship between behavior and response to treatment. One U.S. study (Aumack 1969) and one U.K. study (Rosenthal 1977) were used to provide information on frequency of behavior in psychiatric wards. One U.S. and two U.K. surveys provided information on the opinions of relevant caregivers on what constitutes key behavior. They were Hogarty (1966) (social workers with experience in psychiatric rehabilitation, \( n = 70 \)), Creer and Wing (1974) (relatives living with ex-patients, \( n = 80 \)), and Hall (1977) (psychiatric nurses, \( n = 76 \)).

2. Compiling the data from the source studies to indicate what priority should be assigned to various broad classes of behavior (such as "speech," "recreation," and "washing and hygiene"). This was done by mathematically transforming the ranks ascribed to the different classes of behavior in each study and then combining the ranking scores to produce an overall weighted score for each behavioral class.

In choosing items for the scale, the number of items assigned to each behavioral class was proportional to the weighted scores. The detailed content for each item was reached by further careful analysis of the source studies, within the constraint that every item had to have an equally weighted score (i.e., of equal priority according to the source studies).

This method should ensure that there is no overemphasis or underemphasis on certain behaviors and that the overall scale content reflects the findings of the source studies. Any subsequent combinations of scores, such as factor or total scores, should be meaningful as each item was selected on the basis of an equal weighted score.

**Format.** The design of the format of the scale was particularly influenced by four psychometric texts (Guilford 1954; Oppenheim 1966; Mehrens and Lehmann 1969; Cronbach 1970), and by careful examination of other rating scales (Hall 1980).

Many rating scales use a common item format for all items, even though different types of behavior may require a different item format. In particular, it has been shown repeatedly that deviant or socially embarrassing behavior is not correlated with socially withdrawn behavior (Wing 1961; Baker et al. 1974, 1977; Lewine et al. 1983). While the former type of behavior typically consists of low-frequency, short-duration, actively inappropriate behavior, the latter type typically consists of deficits of behavior that are generally displayed by a patient (e.g., lack of speech). The item format chosen for DB items was therefore different from the format for GB items.

For DB items, a 5-point rating format was initially used, with the response points being: high frequency, high severity; high frequency, low severity; low frequency, high severity; low frequency, low severity; and no occurrence. Items referred to specific weekly or daily frequencies of deviant behavior. The rater ticked a box to indicate the frequency and severity of the observed behavior over the period of 1 week.

For GB items, a standard linear graphic (visual analog) item format was chosen, using a line with three verbal descriptions underneath, one at each end and one in the middle. The descriptions for each item were unidimensional. The reference standard for the "good" description at the right of each line was the standard of behavior appropriate to living in the community, and this reference point was emphasized in the instructions. This clear statement of the assumed norm is an advantage over the frequently used "average"
or "normal" description. The reference point for the "bad" description at the left of each line was the worst level of behavior possible, with the mid-point description representing a level of behavior halfway between acceptable behavior in the community and worst behavior possible. Each description indicates a different level of frequency or quality of performance for the target behavior, and was derived from items in the nine source studies, and from a number of other scales. A guide produced for raters gives expanded definitions about what is meant by each description for each item, as a source of reference. For each item, the rater makes a mark through the line at the point which best represents the observed patient's behavior in the week observed. Figure 1 provides examples of items from the DB and GB sections of the scale as they appear in the final version of the scale.

Experience with the Wing Ward Behavior Rating Scale (Baker et al. 1974, 1977) had demonstrated the tendency for ratings of this type of item to cluster at the "no problem" end; since there is a well-known positioning effect in which raters show a preference for the left-hand side of the scale (Cronbach 1970; Maloney and Ward 1976), the "no problem" category was placed on the right for all items, with the intention of producing a wider distribution of item scores.

Each item was expressed in the form of a question. The response descriptions were worded so as to be simple, to be varied in language, and to refer to actual behavior, not inferred traits or behavior; these are all factors that should contribute toward creating "a good cue" and should improve reliability (Guilford 1954; Jensen and Morris 1960).

One important requirement was to produce a scale that was as short and simple as possible, but also generated the most useful aspects of information. The wording of the scale was made as simple as possible to ensure good "readability" (Flesch 1948; Gunning 1952) by direct-care staff who might not be highly literate. The DB and GB sections of the form itself contained simple instructions for raters on how to fill them in correctly, so that they could complete the ratings without a manual.

The clerical task of completing the scale was made as simple as possible. The rating period was fixed at 1 week, as this is the basic unit of staff allocation of time and also ensured adequate opportunities for observation.

Pilot Studies

The first version of the scale was used in a series of studies with 168 patients, mainly diagnosed as suffering from chronic schizophrenia according to British practice, all under 65 years of age, and all having been continuously in a psychiatric hospital for 2 years or longer (Hall 1978). These studies indicated some essential characteristics of the scale. The most important characteristic was the factor structure of the GB subscale for all 168 patients, on the basis of principal component factor analysis with iteration using varimax orthogonal rotation to yield seven factors.

Subsidiary factor analyses for the six wards, which varied in size between 12 and 45 patients (mean size 28), were also carried out, yielding five factors. Since the wards represented a range of levels of disability, the stability of the factor structure across these six wards supports the factor structure of the following four major factors which appeared in the main analysis: (1) sociability and amount of speech (5 items: factor weights .73 to .82); (2) self-care skills (5 items: factor weights .68 to .83); (3) work (3 items: factor weights .66 to .81); and (4) quality of speech (2 items: factor weights .61 and .69).

For the DB items, only 4 percent of ratings were marked for severity, while 17 percent of ratings were along the frequency continuum, so the severity continuum was eliminated from DB items, producing a 3-point frequency item format. The GB item format was successful in producing item-score distributions and full-scale distributions that used the whole range of the available scores, although item scores showed some evidence of peaking about the end- and mid-cue points.

Although the work factor appeared robust in the factor analyses, experience with ward and most daycare staff showed clearly that there was no opportunity to observe those items adequately for more than a small proportion of patients. Items on the work factor were therefore eliminated, as was one other item on mood that was interpreted too subjectively by different raters.

Subsequently, the work items were developed into a different and separate scale, for use in work settings (Baker 1986). Following this exclusion, and in the light of raters' comments and by reference to the next most important items in the original item analysis, two new items on community skills were then added to the scale (Money Management and Use of Community Facilities).

Included in the pilot studies were two separate validity studies (Hall 1978). Twenty-two patients on a token economy ward were rated on the pilot version of
REHAB and on Wing’s Ward Behavior Scale (the most widely used U.K. rating scale [Hall 1979]), producing a correlation of .72 between REHAB DB and Wing’s Socially Embarrassing Behavior factor, and a correlation of .64 between REHAB GB and Wing’s Social Withdrawal factor. Secondly, a group of 25 patients were assessed with the pilot version of REHAB, and with standardized psychiatric rating scales, psychological tests, and a 40-item checklist of specific items of problem behavior. This multidimensional validity study showed that the assessments relating to ward behavior and psychiatric nursing were measuring an area of functioning statistically independent of the interview-based psychiatric ratings and psychological tests, and were contributing an independent dimension to overall assessment.

The Main Scale

On the basis of these preliminary studies, the final version of the scale was developed. Apart from taking note of the results of the pilot studies—for example, changing the DB items to a 3-point format—attention was paid to redesigning the rating form itself in a way that would make it more acceptable to users. “Readability” scores were calculated to indicate how easily the scale could be understood by raters (Baker and Hall 1980). Phrases and wording were edited to produce unambiguous and short statements. Changes in item content were carefully monitored according to the original data on which content was based. In the final version of the scale, the distribution of item content is close to the original proportions calculated (excluding work-related items). Our clinical experience with the final version of the scale has confirmed a good degree of acceptance by direct-care staff using the scale. Figure 1 contains examples of items from the DB and GB sec-

Figure 1. Examples of Items from the REHAB Scale

Deviant Behavior

2. Was the patient physically violent? 
Violent, (e.g., hit someone, broke something) more than once in the week

7. Did the patient talk or laugh to himself/herself?
Episodes of talking to self, or outbursts of laughing/giggling more than once every day in the week

General Behavior

10. What did the patient do with his/her spare time?

14. How sensible was the patient’s speech?

16. How good were the patient’s table manners?

The reliability of rating scales is ing form itself, and a supplementary “Raters Guide” is provided for reference by the person making the rating.

A standard rater training procedure for raters should further minimize poor reliability. It has been suggested that scales should be easy to administer so that rater training should not be necessary (Lawton et al. 1969), but this seems an unduly optimistic suggestion. While some published rating scales quote reliability where no training was given (Wing 1961), others quote reliability after experience in their use (Cohen et al. 1944).

A standard rater training procedure was accordingly developed for REHAB, and this is fully described in the “User’s Manual” (Baker and Hall 1984). The procedure involves three steps: (1) Two trial patients are identified who are to be observed by the trainee raters for 1 week. (2) The scale administrator meets with the raters and instructs them to rate one of the patients, discussing the ratings, item by item. The trainee raters are then asked to rate the second pre-identified patient independently at some point after the training session. (3) After scoring these ratings, the scale administrator meets with the trainee raters again, and they examine the ratings of the second patient, focusing on agreements and discrepancies. The raters are also given a copy of the “Rater’s Guide” which gives extended definitions of the meaning of each item to help in consistency of interpretation. This training procedure was designed to give the raters understanding of the scale and to identify and correct errors in rating.

Eight different pairs of nurse raters from four long-stay wards who had previously received this standard training program and who worked on the same nursing shift independently rated between 3 and 10 long-stay psychiatric patients each (mean 5.8, median 6). In total, the pairs of raters assessed 47 different patients. For each REHAB item, Spearman rank correlation coefficients have been calculated to show the degree of relationship between the independent ratings. For each item, the correlations are based on the ratings of the 47 patients as completed by the eight pairs of raters. Table 1 shows the correlation coefficients. All the correlations were significant beyond the .001 level. It could be claimed that every item correlates highly with every other item and that these reliability figures are part of a general indiscriminate effect instead of showing specific agreement on specific items. To test this, coefficients were also calculated between every dissimilar item. In the DB section, only one other item correlation was significant at the .001 level, showing the effect to be specific. In the GB section, 118 out of a possible 352 correlations were significant at the .001 level. Correlations were highest for items within the same factor. This pattern of inter-item correlation corresponds closely to the actual inter-item correlation structure of General Behavior (Baker and Hall 1984). Thus, for GB items, although the individual item reliabilities are not completely discrete, they are as would be expected. It should be noted that the correlation coefficients given are for individual items, not, as is usually the case, for overall scores. This is a more stringent test of interrater reliability; each item has a high and statistically significant coefficient.

Other studies have reported on item reliability of the rating scale. Baker (1986) and Bell (1983) report a number of other large samples of independent ratings that corrbo-
Table 1. Spearman correlation coefficients between independent raters (n = 47)

<table>
<thead>
<tr>
<th>Item</th>
<th>r</th>
<th>Item</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Incontinence</td>
<td>.80</td>
<td>13. Initiation of speech</td>
<td>.80</td>
</tr>
<tr>
<td>2. Violence</td>
<td>.65</td>
<td>14. Speech sense</td>
<td>.72</td>
</tr>
<tr>
<td>4. Sexual acts</td>
<td>.67</td>
<td>16. Table manners</td>
<td>.62</td>
</tr>
<tr>
<td>5. Absconding</td>
<td>.61</td>
<td>17. Washing</td>
<td>.63</td>
</tr>
<tr>
<td>6. Verbal aggression</td>
<td>.76</td>
<td>18. Dressing</td>
<td>.72</td>
</tr>
<tr>
<td>7. Talking to self</td>
<td>.88</td>
<td>19. Care of possessions</td>
<td>.87</td>
</tr>
<tr>
<td>8. Mixing on ward</td>
<td>.63</td>
<td>20. Need for prompting</td>
<td>.72</td>
</tr>
<tr>
<td>9. Mixing off ward</td>
<td>.81</td>
<td>21. Use of money</td>
<td>.91</td>
</tr>
<tr>
<td>10. Use of leisure</td>
<td>.62</td>
<td>22. Community facilities</td>
<td>.92</td>
</tr>
<tr>
<td>11. Activity level</td>
<td>.67</td>
<td>23. Overall rating</td>
<td>.65</td>
</tr>
<tr>
<td>12. Amount of speech</td>
<td>.69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N/O = not obtainable.

rate the reliability of the GB section of the scale. Lavender (1984) used the rating scale to assess the effect on patients' behavior of staff changing from nurse uniforms to everyday clothes. Independent raters assessed patients on four separate occasions, twice during baseline and twice after a change to everyday clothes. In total this comprised 80 separate reliability checks. Spearman rank correlation coefficients were significant in all of these reliability checks for DB, and 77 of the checks for GB.

One problem in the DB section of the scale is that it refers to items of behavior that occur sporadically, perhaps only once a week. Since any rater is on a shift and may have days off during the week of rating, it is clear that incidents involving deviant behavior may go unobserved. It would not necessarily be expected that two raters would observe every incident of DB, so this is an inherent source of variability. Indeed this is a problem not specific to REHAB but common to other rating scales that attempt to measure infrequently occurring behavior (Johnson and Bolstad 1973). The use of correlation coefficients to assess the reliability of DB is somewhat limited since the large number of agreements about the nonoccurrence of behavior (zero scores) and, hence, the high level of chance agreement obscures the smaller number of observations of the positive occurrence of DB. This suggests that an alternative method of calculating the reliability of DB items is to use a chance-corrected measure, such as weighted kappa (Hall 1974).

If we set aside the zero scores and look only at observations where DB was actively rated as occurring, we find that 11 percent of the observations of two raters show an important difference. One way to reduce variability of this kind is to introduce a recording system in which each rater must record incidents involving DB on a "master chart" open to all raters for inspection before completion of the rating form. This was used successfully with another rating scale in which DB was featured, and should be applicable to REHAB too (Baker et al. 1974). Also, the REHAB "User's Manual" suggests using a pattern of rating in which two raters on different shifts on 2 consecutive weeks observe and rate patients, and their mean score is used. We consider it important to use some such measure for the DB section of the scale so that the likelihood of observing occurrences of DB is increased. This procedure increases the stability of the score for day-to-day use, but since it eliminates the independence of the raters, it could not be used for formal determination of interrater reliability.

Validity

In the construction of REHAB, items were selected on the basis of three empirical criteria, and the number of items in individual content categories was directly related to the findings of the source studies. Content validity is thus high. Apart from the validity studies quoted with the pilot version of the scale, additional studies have been undertaken with the final version.

One criterion-related validity study compared REHAB with direct observation of patients’ behavior on two long-stay psychiatric wards (Iljon Foreman and Baker 1986). Psychiatric nurses rated 43 patients on REHAB over a 2-week
One of the objectives of REHAB is to assess chronic patients over a range of psychiatric disability. Figure 2 presents information on 577 patients from six different types of psychiatric population (this figure does not include data from any of the 168 pilot study patients). Histograms of total REHAB scores show clear and predictable differences between the six populations, so that, for instance, at the extreme, the scores for day patients living in the community and the scores for patients living in geriatric wards show clearly different distribution patterns.

Another objective of REHAB is to derive scores that can be used to select and discriminate different populations—for instance, to discriminate between patients in the hospital and those in the community. A contrasted groups-comparison method (Anastasi 1959; Maloney and Ward 1976) was used in which the scores of patients in different populations were compared. In a comparison of psychiatric day hospital attenders (mainly psychotic) living in the community ($n = 189$) with patients from “average” long-stay psychiatric wards ($n = 204$), a discriminant function analysis, entering Total DB and each factor score, correctly classified 77 percent of patients by selecting three factor scores (Social Activity, Community Skills, and Self-Care) and the Total DB score.

In a second analysis with the same groups, entering only Total DB and Total GB scores, only Total GB was selected, correctly classifying 75 percent of patients. A similar analysis was conducted comparing the same inpatients from “average” long-stay psychiatric wards ($n = 204$) with psychiatric inpatients selected for a predischarge ward ($n = 57$). Sixty-seven percent of patients were correctly classified by Total GB alone, and

---

**Figure 2. Total General and Deviant Behavior score distributions for different populations**

<table>
<thead>
<tr>
<th>Population</th>
<th>General Behaviour</th>
<th>Deviant Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day Hospital</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 189</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selected for a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>predischarged ward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 57</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate Handicap</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long stay wards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disturbed long</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stay wards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 106</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old long stay wards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geriatric wards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

A psychologist made time-sampled observations of patients over the same 2 weeks using a behavioral coding system devised by McFadyen (1984). The predicted correlations between individual REHAB items and similar types of behavior as directly observed were statistically significant in nearly all cases, whereas REHAB items unrelated to the directly observed behavior were not significantly correlated.
the cutoff score from this analysis was very similar to that of the previous analysis (GB scores of 42 and 40, respectively, out of a total score of 144). A fourth discriminant function analysis, entering Total DB and each factor score, compared the same inpatients from "average" long-stay psychiatric wards (n = 204) with inpatients from disturbed psychiatric wards (n = 83). Seventy-seven percent of patients were correctly classified by the selection of four measures from REHAB (in order of priority; Self-Care, Total DB, Social Activity, and Community Skills). REHAB factor scores are thus able to discriminate between patients at both the least and most handicapped ends of the spectrum of disability on the basis of very few scores, in one case by a single score alone—the Total GB score.

Sensitivity to Change

One stated objective of REHAB is that it should be able to detect changes in patients' behavior. Four controlled studies have been conducted to examine the scale's sensitivity to change. A study was conducted in which 16 long-stay psychiatric patients were exposed to either a community skills program or the standard psychiatric hospital regime, using a matched pairs design in which patients were randomly assigned to experimental or control groups, with premeasurement and postmeasurement (Baker 1986). The experimental group significantly improved in the two factors relevant to treatment—Social Activity and Community Skills as assessed by t tests, although only the Social Activity factor significantly differed from the scores of the control group.

In another study (Tulloch et al. 1986), an "activity therapy" group, derived from the work of Wing (1975), was run with seven long-stay psychiatric patients. Patients were assessed on REHAB every second week by two independent raters over a period of 6 months, which included a baseline and a treatment period. Regression coefficients of score against session number were calculated for each patient on Total DB, Total GB, and each factor; t tests were used to calculate whether the means of regressions coefficients differed from zero. Significant improvement was found in the same two factors as the previous study. A check on the residuals (lag 1 autocorrelation of residuals) showed no significant error effect for the results on these two factors.

The independent investigation of the effects on psychiatric patients of nurses changing from uniform to everyday clothes (Lavender 1984), on the basis of a 4-month baseline and 7-month treatment phase, found that both Total DB and Total GB registered significant changes between the two periods (better with everyday clothes) as assessed by various nonparametric statistical methods. Total GB showed changes from 1 month to another, whereas Total DB only showed changes when data were grouped over several months and compared with grouped data from other months.

A double-blind intersubject crossover drug withdrawal study was conducted over a 12-month period, during which 22 chronic schizophrenic patients living in a hostel were assessed on REHAB every month (Mathur and Hall 1981; with further analysis in Baker 1986). Patients' scores on REHAB were regressed on month number, with t tests used to assess the significance of the regression coefficient. As figure 3 demonstrates, all patients showed a significant trend toward deterioration on Total DB. On Total GB, however, only the patients who later relapsed (as determined by the nursing staff independently of the REHAB scores) showed a significant trend toward deterioration (p < .001), which existed before relapse and was not shown by the other group.

Further information on each of these studies can be found in Baker (1986).

Overall, the change in scores in the expected direction, the existence of clear trends in the repeated measures studies, and the fact that control groups/periods did not show improvement provide reasonable evidence that the GB section of the scale is a sensitive measure of a patient's behavioral change.

Factorial Structure

As previously mentioned, various factor analyses were carried out during the pilot stage of REHAB's development (Hall 1978). Later, during the main stage of the study in which REHAB was used in (almost) its present form, further data from 821 psychiatric patients were gathered (Baker and Hall 1984). This provided information from several different samples, which were each separately factor analyzed, independent of the pilot study analyses. Information about the final factor structure for REHAB was therefore based on a variety of different analyses.

The type of factoring used was principal component factor analysis with iteration, employing varimax orthogonal rotation (Nie et al. 1975). Table 2 shows the varimax-rotated factor matrix for 508 patients from 18 long-stay psychiatric wards. The table shows the factors...
Figure 3. Repeated measurement during drug withdrawal: Group means

Table 2. Varimax-rotated factor matrix for long-stay populations (n = 508)

<table>
<thead>
<tr>
<th>Items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Mixing on ward</td>
<td>.22</td>
<td>.71</td>
<td>.22</td>
<td>.21</td>
<td>.06</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>9. Mixing off ward</td>
<td>.10</td>
<td>.73</td>
<td>.15</td>
<td>.12</td>
<td>.31</td>
<td>.01</td>
<td>-.01</td>
</tr>
<tr>
<td>10. Use of leisure</td>
<td>.24</td>
<td>.78</td>
<td>.15</td>
<td>.12</td>
<td>.11</td>
<td>.08</td>
<td>.05</td>
</tr>
<tr>
<td>11. Activity level</td>
<td>.24</td>
<td>.51</td>
<td>.06</td>
<td>.30</td>
<td>.21</td>
<td>.11</td>
<td>.07</td>
</tr>
<tr>
<td>12. Amount of speech</td>
<td>.24</td>
<td>.45</td>
<td>.34</td>
<td>.74</td>
<td>.16</td>
<td>.03</td>
<td>.05</td>
</tr>
<tr>
<td>13. Initiation of speech</td>
<td>.20</td>
<td>.54</td>
<td>.27</td>
<td>.57</td>
<td>.12</td>
<td>.06</td>
<td>-.02</td>
</tr>
<tr>
<td>14. Speech sense</td>
<td>.30</td>
<td>.28</td>
<td>.66</td>
<td>.10</td>
<td>.15</td>
<td>.06</td>
<td>.11</td>
</tr>
<tr>
<td>15. Speech clarity</td>
<td>.30</td>
<td>.17</td>
<td>.70</td>
<td>.30</td>
<td>.10</td>
<td>.03</td>
<td>-.07</td>
</tr>
<tr>
<td>16. Table manners</td>
<td>.71</td>
<td>.18</td>
<td>.22</td>
<td>.12</td>
<td>.11</td>
<td>.00</td>
<td>-.05</td>
</tr>
<tr>
<td>17. Washing</td>
<td>.85</td>
<td>.21</td>
<td>.18</td>
<td>.14</td>
<td>.06</td>
<td>.04</td>
<td>-.01</td>
</tr>
<tr>
<td>18. Dressing</td>
<td>.88</td>
<td>.15</td>
<td>.16</td>
<td>.10</td>
<td>.12</td>
<td>.03</td>
<td>-.08</td>
</tr>
<tr>
<td>19. Care of possessions</td>
<td>.76</td>
<td>.14</td>
<td>.16</td>
<td>.08</td>
<td>.14</td>
<td>.12</td>
<td>.07</td>
</tr>
<tr>
<td>20. Need for prompting</td>
<td>.75</td>
<td>.29</td>
<td>.18</td>
<td>.17</td>
<td>.16</td>
<td>.02</td>
<td>.34</td>
</tr>
<tr>
<td>21. Use of money</td>
<td>.41</td>
<td>.36</td>
<td>.33</td>
<td>.13</td>
<td>.42</td>
<td>.13</td>
<td>.10</td>
</tr>
<tr>
<td>22. Community facilities</td>
<td>.22</td>
<td>.42</td>
<td>.16</td>
<td>.15</td>
<td>.67</td>
<td>.05</td>
<td>.00</td>
</tr>
<tr>
<td>23. Overall rating</td>
<td>.49</td>
<td>.35</td>
<td>.32</td>
<td>.17</td>
<td>.33</td>
<td>.45</td>
<td>.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factors</th>
<th>Self-Care</th>
<th>Social Activity</th>
<th>Disturbed Speech</th>
<th>Speech Skills</th>
<th>Community Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name given to factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The item weights within the chosen factors are high, especially when one considers that weights over .2 are statistically significant. There is for most factors a reasonable separation in item weights between items included in a factor and those excluded.

Separate factor analyses are reported in the "User's Manual" for a Scottish sample (n = 367), a Welsh sample (n = 167), and an English sample (n = 164) (Baker and Hall 1984). These samples control to some extent for any spurious factoring of items due to their order and layout on the assessment form, as the order and layout differed somewhat between the three samples while the content of questions was the same. The three analyses show the same factors emerging as in table 2 in virtually the same order with similar weights and similar separation of factors from excluded items, with only one exception of any note—one sample excludes items 11, 12, and 13 from the Social Activity factor. The consistency between the separate samples is such
as to corroborate the chosen factor-analytical structure.

Discussion

A new scale has been designed to assess the behavior of people with chronic or severe psychiatric disabilities. It can be used in a variety of settings. Because ratings of abilities. It can be used in a variety of settings. Because ratings of abilities. It can be used in a variety of settings. Because ratings of abilities. It can be used in a variety of settings. Because ratings of abilities. It can be used in a variety of settings. Because ratings of abilities. It can be used in a variety of settings. Because ratings of abilities. It can be used in a variety of settings.

Apart from its use as a research instrument, REHAB has been designed to be as clinically relevant as possible, in two ways in particular: (1) The scale is presented in the form of an integrated package of materials, which take the professional user through all the steps necessary in using, scoring, interpreting, and presenting findings. (2) As services for chronic psychiatric patients vary considerably, the scale is part of a flexible package capable of use in different settings and according to different assessment purposes.

The “User’s Manual” explains how to use and interpret REHAB findings according to seven different assessment aims as follows: establishing a baseline, assessing change, selecting patients with potential for discharge from hospital, selecting disturbed/severely handicapped patients, selecting groups of patients, planning treatment/intervention for individual patients, and lastly planning treatment, intervention, or reorganization for wards, groups of wards, or institutions. The “User’s Manual” also contains tables of norms that enable the clinician to compare an individual patient with populations of patients of similar age or length of hospital stay, and with other populations, such as “clients living in the community,” “hospitalized long-stay patients,” or “disturbed patients.” Probability tables are derived from the previously mentioned discriminant function analyses also provide guidelines for the clinician about the probability of the patient’s membership in different groups, such as “living in the community” versus “living in hospital,” based on REHAB scores. These probabilities should be useful in indicating how closely the patient resembles those surviving in the community as opposed to those living in psychiatric hospital.

REHAB has been used in a variety of ways. Hall incorporated it into a computer-based case register which enabled all long-stay patients in a Welsh psychiatric hospital to be called up routinely for regular review (Trickey 1979). In Scotland, seven clinical psychologists used REHAB to assess ward populations, and then based the planning of ward reorganization or rehabilitative programs on the REHAB findings (Baker 1986). It has been used in various surveys such as the assessment of a day hospital population (Pryce et al. 1983) and an industrial therapy department (Wojciechowski 1984), and the identification of patients in a long-stay hospital who might be suitable for transfer to the community services (Flett et al. 1984). Psychiatric nurses have used it to help revise the treatment program of a ward for young chronic psychiatric patients (Barker 1985), to measure improvement in an individual patient (Bluteau and Long 1986), to plan and evaluate day care for psychiatric patients (Greig et al. 1985), and to restructure the operation of a long-stay psychiatric hospital (Bell 1985). It has also been modified for use with patients’ relatives (Hewitt 1983).

In 1984 the Personal Social Services Research Unit of the University of Kent was commissioned by the British Department of Health and Social Security to monitor and evaluate a number of rehabilitation projects in England. In their initial survey (Personal Social Services Research Unit 1985), they found that some of the mental illness projects were using REHAB and had found it useful in aiding decisions about selection of patients for rehabilitation programs or discharge to the community. The au...
thors hope that REHAB will prove to be a practical instrument that can aid both clinicians and researchers in their work with psychiatrically disabled patients.

References


Baker, R.D. “A Methodology for Selecting Items for Assessment Scales.” Department of Psychology, Royal Cornhill Hospital, in preparation.


Baker, R.D., and Hall, J.N.


Baker, R.D.; Hall, J.N.; and Hutchinson, K. A token economy project with chronic schizophrenic pa-


Flett, S.; Jude, P.; Maltby, I.; and Thomas, J. Interim report of the community care working party, Mental Illness Unit, to the South Lincolnshire Health Authority. Psychology Department, Rauceby Hospital, Sleaford, Lincolnshire, 1984.


Hall, J.N. Inter-rater reliability of ward rating scales. British Journal of


Oppenheim, A.N. Questionnaire De-


Personal Social Services Research Unit. Care in the Community: Assessment, Selection and Rehabilitation. PSSRU, University of Kent, Canterbury, Summer 1985.


Trickey, A.J. "An Information Coordinating Centre for Mental Health: Report to South Glamorgan Health Authority." Computer Coordinating Centre, South Glamorgan Health Authority, Cardiff, Wales, 1979.


Acknowledgments

The early and later stages of the development of the scale were supported, respectively, by grants from the Department of Health and Social Security and the Grampian Health Board. Leeds University, Aberdeen University, and South Glamorgan Health Authority provided computer facilities. Many psychological, psychiatric, and nursing colleagues, especially Sarah Pickering, provided great help.

The Authors

Roger Baker, Ph.D., is Area Research Coordinator, Department of Psychology, Royal Cornhill Hospital, and Clinical Lecturer, Department of Mental Health, Aberdeen University, Aberdeen, Scotland. John N. Hall, Ph.D., is District Clinical Psychologist, Department of Psychology, Warneford Hospital, and Clinical Lecturer in Clinical Psychology, Department of Psychiatry, Oxford University, Oxford, England.
Appendix: Layout of REHAB

The questions used for each item are grouped according to the factor structure.

- Scale title and logo
- Basic data about patient: name, date of birth, etc.
- General instructions about both parts of the scale
- Deviant Behavior section heading
- Specific instructions for the Deviant Behavior section of the scale

1. Was the patient incontinent?
2. Was the patient physically violent?
3. Did the patient hurt or mutilate himself/herself?
4. Was the patient sexually offensive in any way?
5. Did the patient leave the ward or hospital without arrangement?
6. Did the patient shout or swear at others?
7. Did the patient talk or laugh to himself/herself?

- An item concerned with deviant behavior in the last year
- General Behavior section heading
- Specific instructions for the General Behavior section of the scale

Social Activity factor
8. How well did the patient get on with others on the ward or unit?
9. How much did the patient mix with others off the ward or unit?
10. What did the patient do with his/her spare time?

Disturbed Speech factor
11. How active was the patient?
12. How many words did the patient use when he/she spoke?
13. How much did the patient initiate conversation?
   (12 and 13 also make up the Speech Skills factor.)

Self-Care factor
14. How sensible was the patient’s speech?
15. How clearly did the patient speak?

- An item concerning how typical the patient’s behavior was in the last week.
- An “any other comments” item.