

## EDITORIAL

# The role of rating scales in psychiatry

The first application of rating scales in psychiatry goes back to soon after the First World War, but the real development did not occur until after the Second World War. Their use has increased so much that it is scarcely possible to look through any copy of a general psychiatric journal without finding at least one paper which involves the use of a scale of some sort. Such papers are always concerned with research; up till now the only attempt to use rating scales in clinical practice has been related to the efforts made to store case records on computers.

The term 'rating scale' was originally used to define a series of items which quantified or placed in rank order, the manifestations of a single variable, e.g. aggressiveness. Another form of rating scale was applied to individuals to quantify the extent to which they possessed a given attribute, e.g. persons could be rated on their aggressiveness. For most purposes, especially clinical, the term is often used to describe a set of scales which have some intrinsic relationship to each other. The individual scales are then referred to as items of the total scale. There are many kinds of scales in current use and they can be classified in different ways. The most obvious is in terms of the user: self-rating scales and those used by an observer, who can be skilled, e.g. psychiatrist or psychologist, or semi-skilled such as a nursing aid, or unskilled, e.g. the relative of a patient. The data of the first type can be obtained from structured, semi-structured or free interviews. Classification can be according to the form of the items: graded items record degrees of severity or relevance; in check-lists the items are scored as present or absent; and there are forced-choice items in which the rater has to choose which of two alternatives is most applicable. Scales can also be considered in relation to their content, e.g. symptoms, behaviour in the ward, social adjustment, family relations, functional capacity in an occupational therapy or an industrial therapy setting. Finally, and most important, scales can be classified according to their function. There are four of these: (1) intensity scales which measure severity of illness and also response to treatment; (2) prognostic scales, including prediction of response to treatment; (3) scales for selection of treatment by means of differential indicators; and, finally, (4) scales for diagnosis and classification.

A number of misconceptions about rating scales must first be considered. A rating scale is only a particular device for recording information about a patient. The data provided by it are therefore no better than their basis. Inadequate, misleading or incorrect information is not improved by recording it on a rating scale. For clinical purposes, the best way of describing a patient is by a free and full psychiatric case history. When this is reduced to a rating scale, much information is lost. For some purposes the loss may be serious, but in appropriate circumstances it may be of no account. In a sense, rating a patient is fitting him into a Procrustean bed; anything which does not conform to the requirement of the scale has to be ignored, i.e. deleted. For example, the rating of a schizophrenic patient on a scale for depression would certainly produce a set of scores and it is conceivable that they might have some sort of meaning, but they do not give an adequate description of the patient's condition. The reverse also occurs. The presence of a large number of irrelevant items in a scale encourages raters to attempt to fill in at least some. Although the information provided by a rating scale is limited, it is valuable because it is uniform for all patients and all occasions, and it is standard in its significance because the items, their grades and manner of use have been previously defined. Rating scales therefore permit comparison between different patients and between different occasions for the same patients. They do this with adequate reliability and validity, which is more than can be said for 'free' case histories and diagnostic labels.

Rating scales are not really suitable for exploring a new field of knowledge. Their construction requires much practical experience and an appropriate body of theory; in a sense they are an

end-product. A scale for grading symptoms depends upon the extensive clinical experience which has defined those symptoms. For example, in order to construct a scale for measuring the efficiency in working capacity of chronic schizophrenics in an industrial therapy unit and the response to treatment, it is first necessary to determine in what way such patients are deficient as compared with normal workers. An appropriate background of clinical theory is also necessary. What is recorded on a scale will consist either of direct observations or of inferences from them, though in practice it is not possible to make an absolute distinction between these two. When the scale is concerned with directly observable behaviour, including the statements of the patient, then what is included in the scale will be determined by what is regarded as relevant on the basis of theoretical considerations. When the behaviour is interpreted, then clearly the interpretations depend upon the theoretical background used in the construction of the scale or by a rater. This applies even to the lowest level of interpretation, in which the behaviour is considered in relation to its setting. The first step in the preparation of a scale therefore consists of the identification of the 'field of discourse', not the other way round. Thus a scale for personality may be based on the notion of traits, on construct theory or on psychodynamics. In the same way, a psychiatric rating scale could be concerned with symptoms, personal relations or unconscious mechanisms. A scale could be devised for the measurement of the process of 'individuation' in a patient undergoing analytical psychotherapy, but it is unlikely that it would wean non-Jungians away from their own theories. However, if repeated attempts to construct such a scale failed to show reliability or validity, the underlying theory would come to be regarded as unsatisfactory and requiring reformulation, at the least.

The most frequent use of rating scales in psychiatry, and therefore the most important, is for assessment of improvement in clinical trials. Their use for this purpose is so common that some clinicians appear to believe that without a rating scale a trial is invalid or 'not scientific'. This is putting the cart before the horse. If in a particular trial the patients included are unsuitable, the amount of treatment inadequate, time of observation too short (or too long) or there are too many drop-outs, then no conclusion can be drawn from the trial, however many rating scales are used and however elaborate the statistical analysis may be. It may be added that even if all these are satisfactory, but the type of patient and mode of selection not clearly described, it may be impossible to know what are the circumstances to which the results of the trial are applicable and therefore it is still of no practical value.

The traditional methods of clinical assessment of improvement, in terms of 'no improvement', 'mild', 'moderate', 'good' and 'complete' recovery (and also grades of worsening) are as good as ever, and indeed rating scales are based on them. The difficulty is that these grades are extremely difficult to define and are lacking in detailed information. This is the chief reason for the popularity of rating scales in clinical trials. It is also a great convenience that the data from rating scales can be given in numerical form and, under appropriate restrictions, these numbers can be submitted to statistical analysis as if they were true measurements on an interval or ratio scale. It is convenient but not essential. The effort and time required to plan and conduct a clinical trial is so great that the saving of statisticians' time is a relatively trivial matter.

The extensive development of the use of rating scales in psychiatry has been, in a sense, a pioneering effort. The value of rating scales and an understanding of the theory of the techniques in their construction and use are now beginning to penetrate other fields in general medicine. At the very least, the use of such scales, by directing attention to the importance of subjective well-being, may inhibit the introduction of therapeutic procedures where 'the treatment is worse than the disease'.

In the last 30 years the variety of treatments available to psychiatrists has greatly increased. We now have, in addition to several different classes of drugs, ECT, individual and group psychotherapy of various kinds, behaviour therapy, social therapy and rehabilitation and even the possibility of bio-feedback. Since all of them are limited in their efficacy we would like to know if they are supplementary or not, and it would be desirable to know beforehand which treatments would give the best results for a given patient or, in other words, how to allocate the patients to different treatments. These are the functions of prognostic and selection scales. The deficiencies of traditional clinical

approaches, e.g. diagnostic categories, are clear, so there is room for improvement by using the detailed and standard information provided by scales.

Such scales differ from those used for assessment of severity of illness and response to treatment in the nature of their content. A scale for measuring severity of illness consists of items which are related to severity and which are sensitive to changes in the patient's condition. A predictive scale can include 'fixed' items such as family history, number of previous attacks, length of illness, and type of personality. It is therefore not a 'rating' scale in the strict sense, but the principles of design and methods of use are the same. A number of such scales have been developed but none has yet come into extensive use in ordinary clinical practice. There are many reasons why this should be so, but one of them is that the scales are not yet good enough.

Rating scales can also be used as an aid to the solution of theoretical problems concerning the classification of mental illnesses. Although interest in this field has been increasing only in the last few years, one of the pioneer uses of rating scales was concerned with this problem. Nearly 50 years ago, T. V. Moore devised a rating scale for the assessment of symptoms of schizophrenia and examined the intercorrelations of the items by a primitive form of factor analysis akin to cluster analysis. He was able to identify the four classical syndromes of schizophrenia: simple, hebephrenic, catatonic and paranoid, and identified also a fifth syndrome. The methods in current use are much more sophisticated and have yielded valuable results. For example, in the classification of depressions and in the distinction between anxiety states and depressive disorders. Nevertheless, different investigators have found opposite results and have come to contradictory conclusions. Factor analysis is now supplemented by different methods of cluster analysis which have become practicable because of the availability of large computers. The future will see much work on these lines and the result will be fruitful. Nevertheless, too much must not be expected from these techniques. They have been applied in the field of cognitive psychology for over half a century and controversy about their results is as lively as ever.

MAX HAMILTON

#### REFERENCES

- Buros, O. K. (1972). *Seventh Mental Measurements Yearbook*. Gryphon Press: Highland Park, N.J.
- Goldberg, L. R. (1974). Objective diagnostic tests and measures. *Annual Review of Psychology* 25, 343-366.
- Guilford, J. P. (1954). *Psychometric Methods*. McGraw-Hill: New York.
- Pichot, P. & Olivier-Martin, R. (eds.) (1974). *Psychological Measurements in Psychopharmacology*. Karger: Basle.