Brit. J. Psychiat. (1981), **138**, 512–514

Comments

RATING SCALES

One of the major impediments to scientific progress in psychiatry is the confusion which surrounds the instruments with which the severity of disorders are assessed. It is certainly true that large numbers of measurement scales are published and an even larger number of *ad hoc* scales are used by individual researchers in their particular studies. However, it is very rare for any further work to be undertaken in an attempt to improve a scale once it has been published, and work to compare the merits and drawbacks of various scales is an even rarer event. Since the choice of a rating scale is often an arbitrary decision (Carroll *et al*, 1973), undertaken without much thought concerning the characteristics of the scale or whether it is best suited to the study, it is not surprising that researchers using rating scales repeatedly report contradictory findings; in fact the surprising thing is that agreement between one study and another ever occurs.

Questionnaires and scales have three main uses in psychiatric research: as screening devices to detect the incidence of a disorder in a community; as instruments to establish the pattern of symptoms or other characteristics in an individual or in a group of patients; and as rating scales or measures of the severity of disorders. This paper is concerned only with the last of these uses.

Rating scales may be designed for administration by a trained observer who may make his observations and complete the scale during an interview or in the setting of daily contact with the patient; or scales may be designed to be completed by the patient. Both types of scale have their uses and also their drawbacks. The deficiencies of self-assessment scales are very obvious; they can only be used by cooperative patients who are also literate and not too ill; they must also not suffer from a condition where non-comprehension (e.g. senile dementia) or falsification (e.g. anorexia nervosa) of the scale responses would be likely to occur. The deficiencies of observer scales are less obvious but just as serious if not recognized and allowance made. The main drawback is that of rater bias, that is the user is influenced in his scoring by a general expectation of how ill the patient 'ought' to be; for instance it is generally assumed that patients are more ill before they commence on a drug trial than they are at the end of it. A second drawback is that raters are influenced in their scoring of severity by their general experience of patients suffering from the disorder but whereas one rater's experience may be confined to mild cases in the community another may have a long acquaintence with severe cases in hospital.

Carroll and his colleagues (*op. cit.*) have delineated the properties of rating scales under the following headings: orientation, sensitivity, information access, utility and specificity. *Orientation* is the bias inherent in the items which compose the scale; this aspect of five ward rating scales for use in hospitalized psychotic patients was studied by Hall (1977); he noted that, although these scales may all have originally been introduced for a specific purpose, they tended to be used interchangeably or in the usual arbitrary way. Nonetheless the orientation varied considerably; for example in one scale 40 per cent of the content was concerned with speech disorder and none with aggressive behaviour whereas in another scale 25 per cent of the content was concerned with aggression and only 12 per cent with speech disorder. Another example of difference in orientation is that of two observer scales for depression: whereas in the Hamilton Rating Scale (Hamilton, 1967) there is a considerable emphasis on somatic symptoms and observed psychomotor behaviour, these elements are missing from the Montgomery-Åsberg Depression Rating Scale (1979) in which the emphasis is upon the patient's report of his mood.

The *sensitivity* of a scale refers to its ability to distinguish in numerically significant terms between various degrees of severity of illness throughout the whole range for that disorder i.e. from very severe to symptom free. For instance if a scale distinguishes clearly between mild and moderate degrees, between moderate and severe degrees but not between severe and very severe degrees of the disorder, then it is not sensitive when used in very severe degrees of the disorder. *Information access* and *utility* have related meanings. The former term applies to aspects of an illness which the scale cannot assess; for instance self-assessment scales cannot probe the vast range of somatic symptoms which may occur in the setting of an anxiety state or delusional denial of illness which may accompany depressive illness. The *utility* of the scale refers to its ease of use by the patients for which

it is designed; both observer and self-assessment scales which pose obscure concepts in long-winded phrases will suffer from a defect in their utility.

The *specificity* of a scale is the degree to which high scores are achieved in the disorder for which it was designed and low scores in all other disorders. However a lack of specificity is inherent in the nature of psychiatric disorders which are rarely composed of non-overlapping symptoms; for instance depression is common in anxiety neurosis and, recognizing this, depression items appear in the Hamilton Anxiety Scale (Hamilton, 1959); therefore depressed as well as anxious patients might achieve a high score on the scale but this fact does not invalidate its use as a measure of the severity of anxiety neurosis.

Most rating scales for psychiatric disorders are composed of items representing the prominent symptoms of the disorder to which the rater ascribes numerical values indicating the intensity of the individual items; the scale score is the sum of the item scores. This procedure has been criticized and it has been pointed out that it is illogical to add together scores of two such different symptoms as anorexia and insomnia. However the procedure has been defended on mathematical grounds by Hamilton (1968) and, in practice, the method probably works better than scoring a single symptom, say depressed mood in depressive illness, and representing the severity of the disorder by that score. A good rating scale should consist of not too many items and not too few and they should have been selected by both the intuition of clinical experience and by the statistical process of item analysis but unfortunately these desiderata apply to very few scales.

Another form of scale is generally referred to as a global severity scale; the global rating is the numerical value allotted by the rater after a short, or longer, acquaintance with the patient and represents his general impression of how ill he considers the patient to be. Visual analogue scales have been developed by Aitken (1969) and further validated by Luria (1975). In this form of rating the patient makes a mark on a 10 cm line, the ends of which are anchored by definitions e.g. 'as depressed as possible' or 'as depressed as I have ever been' to 'normal mood'. A major drawback of this procedure is that the patient has only his own experience by which to judge the severity of his disorder. Moreover very careful choice of words are necessary, especially in the area of mood, for, as Leff (1978) has pointed out, patients and psychiatrists often mean different things by the use of such words as depression. Nonetheless the visual analogue method probably has a useful role in the assessment of the progress of an individual patient.

An interesting development in the measurement of symptoms is the Personal Questionnaire Rapid Scaling Technique (PQRST) developed by Mulhall (1976, 1978). This is an ingenious and useful technique in which the therapist and patient work out, before treatment commences, which symptoms or areas of discomfort will be used to assess progress. The device has an in-built reliability and is not governed by the usual rigidities of the rating scale; for instance such symptoms as pain in the jaw, arousal during sexual intercourse and anxiety when the spouse is late returning home may be the chosen items on which progress is assessed (the scores on each item are considered separately and not added together as in the usual form of rating scale). A drawback is the difficulty in replication of the findings by other workers unless the precise words in which the symptom is formulated are known.

It is unusual for others workers to attempt the analysis of the items of a published scale. However both the Hamilton and the Beck depression scales were subjected to a careful item analysis by Bech *et al* (1975). They found that subscales of both scales led to an improvement in sensitivity. Another example of modification of an existing scale was undertaken by the author and his colleagues (Snaith *et al*, 1976, 1978); the first of these studies led to a development of a depression self-assessment scale into four subscales relating to depression and anxiety; in the second paper the common mood disorder of irritability was added to the other moods. Improvement may come about through a new approach to the use of an existing scale as was the case when Robertson and Mulhall (1979) applied a grid scoring method to an obsessional scale, the Leyton Obsessional Inventory.

The time has probably now arrived for a moratorium to be declared upon the publication of new scales and for researchers to spend their effort in taking a careful look at existing scales to see how they may be improved and to compare their properties. It is certain that researchers in psychiatry and allied sciences stand in much need of good guidance based upon careful work. Until more information is available about the properties of various scales and their suitability for use in various populations and types of disorder, assessment of severity should be based upon two or more different types of measure such as a global scale completed by nurses and an observer scale completed by the researcher, or by an observer scale and a self-assessment scale.

R. P. SNAITH

References

AITKEN, R. C. B. (1969) Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society of Medicine*, **62**, 989–96.

BECH, P., GRAM, L. F., DEIN, E., JACOBSEN, O., VITGER, J. & BOLWIG, T. G. (1975) Quantitative rating of depressive states. *Acta Psychiatrica Scandinavica*, **51**, 161–70.

CARROLL, B. J., FIELDING, J. M. & BLASHKI, T. G. (1973) Depression rating scales. *Archives of General Psychiatry*, **28**, 361–6.

HALL, J. N. (1977) The content of ward rating scales for long-stay patients. *British Journal of Psychiatry*, **130**, 287–93.

HAMILTON, M. (1959) The assessment of anxiety states by rating. *The British Journal of Medical Psychology*, **32**, 50–5.

—— (1967) Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, **6**, 278–96.

—— (1968) Some notes on rating scales. *The Statistician*, **18**, 11–17.

LEFF, J. P. (1978) Psychiatrists' versus patients concepts of unpleasant emotions. *British Journal of Psychiatry*, **133**, 306–13.

LURIA, R. E. (1975) The validity and reliability of a visual analogue mood scale. *Journal of Psychiatric Research*, **12**, 51–7.

MONTGOMERY, S. A. & ÅSBERG, M. (1979) A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, **134**, 382–9.

MULHALL, D. J. (1976) Systematic self-assessment by PQRST (Personal Questionnaire Rapid Scaling Technique). *Psychological Medicine*, **6**, 591–8.

—— (1978) *The Manual for Personal Questionnaire Rapid Scaling Technique*. Windsor: National Foundation of Educational Research Publishing.

ROBERTSON, J. R. & MULHALL, D. J. (1979) The clinical evaluation of obsessionality: a development of the Leyton Obsessional Inventory. *Psychological Medicine*, **9**, 147–55.

SNAITH, R. P., BRIDGE, G. W. K. & HAMILTON, M. (1976) The Leeds scales for the assessment of anxiety and depression. *The British Journal of Psychiatry*, **128**, 156–65.

—— CONSTANTOPOULOS, A. A., JARDINE, M. Y. & MCGUFFIN, P. (1978) A clinical scale for the self-assessment of irritability. *The British Journal of Psychiatry*, **132**, 164–71.

The author wishes also to draw attention to a further recent paper which examines the criteria for rating scales:

HALL, J. N. (1980) Ward rating scales for long stay patients: a review. *Psychological Medicine*, **10**, 277–88.