Introduction

Peter Tyrer and Caroline Methuen

Many years ago one of us (P.T.) was lecturing on the use of rating scales at a conference on research methodology. In a light-hearted way, the well-known acronym for the standard format of a rating scale, SPITZER, was introduced. 'Of course', went the explanation, 'we all know that the initials of the name stand for "Structured Psychiatric Interviews To Zealously Enhance Research", and the core of research methodology is to remember this, over and over again, when carrying out your research'. This explanation was a little too convincing because subsequently several people commented that they did not realise that Spitzer was only an acronym, not a real person, and it was an eye-opener to understand the real meaning of the word. We hasten to remind you that Robert Spitzer is a real person, who has added a great deal to the science of rating scales, and is mentioned several times in this booklet. One of Bob Spitzer's famous saws is, 'if it exists, it can be measured', and these seven words offer both a rationale and a strategy for using such scales. In this booklet we have unashamedly gone for a measure of esteem that many find intensely irritating, the citation rate, because we feel that the more a scale is cited the more value it is to the researcher, and particularly the systematic reviewer. Such a reviewer measuring temperature can accommodate the Fahrenheit and Celsius scales, but would be very put out if there were forty other scales also measuring temperature in completely different ways. By giving the citation rates (as of 2006) of each scale we are not necessarily saying the most cited one is the best, but, other things being equal, if most investigators chose a scale that is very widely used it would be much appreciated by the reviewer and ultimately by the researcher too. Nevertheless, the many-faceted presentations of psychiatry mean that often a standard scale is not appropriate for the subject matter and so a much less frequently cited scale would be better in a particular project. So the exposition of several scales is sometimes necessary in order to achieve the best fit, and very occasionally it may be necessary to construct your own scale for a specific piece of research: as we make very clear, this should be done only as a last resort.

So it only remains for you to look at the menu, ask the waiter and, if needed, the cook – don't be afraid to write to the author of the scale – to find out the exact nature of the fare, and then make your choice. *Bon appétit*.

Rating scales in psychiatry†

Peter Tyrer and Caroline Methuen

One of the most difficult tasks for aspiring research workers is choosing a rating scale. In an ideal world this should not be a difficult decision. Certain problems require special evaluation and, provided the problem has been recognised before, a suitable rating scale will exist for the purpose. If the rating scale is well established and is clearly the leader in the area, it will choose itself and there should be clear instructions on what training and expertise the researcher will need before the scale (or questionnaire) is applied. However, in practice choosing a rating scale is seldom this straightforward. This is mainly because there are too many rating scales and it is extremely difficult for the novice, and often even the expert, to choose the right scale easily. The rating scales described here are only a selection from a much larger pool; the abundance of new scales has made it impossible to cover the territory adequately. This booklet is therefore a general guide which should enable the researcher to identify the most appropriate scales for their area of interest, but a little more research will be required before the final choice of a scale is made. Hence we have given the main references for a large number of scales in the absence of space for an adequate review of each, on the premise that the wider the pool the better the eventual selection.

Choosing a rating scale

Figure 1 indicates the bumpy journey that the researcher will have to take before feeling confident that the right instrument has been chosen for the problem under investigation. The scales published in this booklet are by no means exhaustive so do not feel that it is unjustified for you to use a scale of your own choosing if you cannot find a measure for the subject under review in the pages below.

†This was first published as chapter 11 in Freeman, C. & Tyrer, P. (2006) *Research Methods in Psychiatry* (3rd edn). Gaskell.

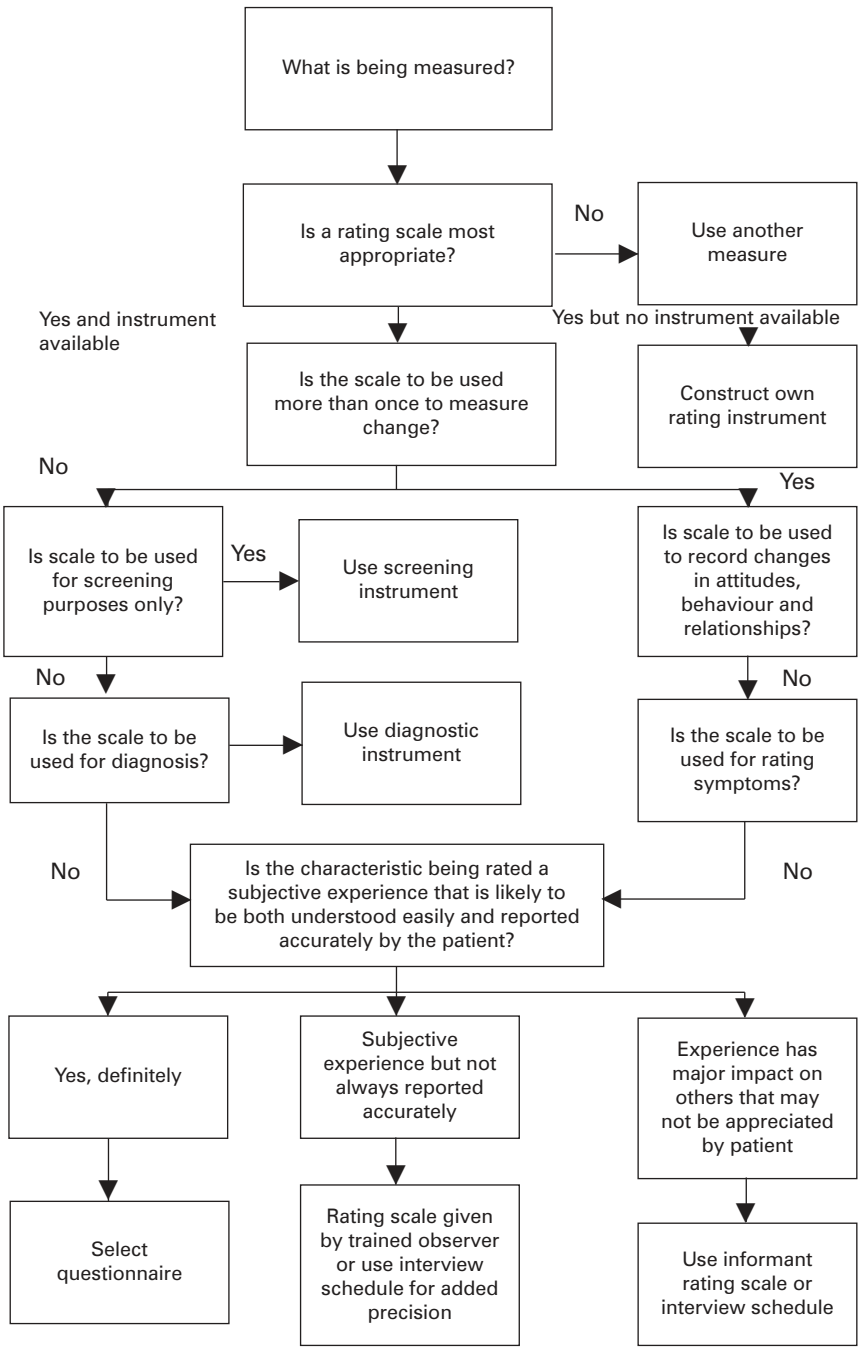

Fig. 1 Flow chart for selection of rating instruments in psychiatric research.

What is being measured?

Rating scales may not always be necessary in a study. As the use of rating scales involves administering an interview to an individual (patient or informant), the procedure is liable to natural attrition in any study, ranging from refusal to take part through to inability to follow-up. However, some other measures (e.g. admission to hospital) can be obtained from other sources and are more likely to yield complete data. There is also a strong and unnecessary tendency for junior researchers to collect as much data as possible without regard to its purpose. Investigators should elect at the design stage to ask whether every single item of information is essential, with the objective of eliminating at least half. The main advantage of simpler methods using few variables is that larger numbers of patients can often be accessed and so more robust findings are likely to emerge. It is therefore reasonable for the researcher to ask the question 'Can I get away without using a rating scale in this project?' It will save a tremendous amount of time and trouble if rating scales and questionnaires are avoided.

What is being measured and why?

There are three main uses for rating scales in psychiatry. The first is as a screening instrument which identifies a population with the condition of interest but could include some people without the condition. A screening instrument should have high sensitivity even though this may be achieved at the expense of low specificity.

The second reason for using a rating scale is to identify a feature that is felt to be important. Quite often this is a psychiatric diagnosis, but it could be any characteristic. The point of using the rating scale is to more accurately measure this characteristic and thereby improve the quality of the research, and also to compare the findings with other studies. First, for example, if one wanted to assess whether a specific personality disorder was associated with childhood physical abuse, the researcher might consider it necessary to assess such abuse (e.g. using the Child Trauma Questionnaire) rather than simply asking the patient a yes/no question.

The third reason for using an instrument is to record change, either spontaneously or following some type of intervention. This raises several other important questions. Is the instrument easy to administer, does repeated assessment lead to practice effects and is the administration of the instrument prone to bias of any sort?

The answer to these questions should determine the nature of the rating instrument selected and whether it is to be self-administered (i.e. a questionnaire) or administered by another person such as a researcher.

Source of information

Reliability always tends to increase with more structured scales and with trained interviewers. There is an understandable tendency to select such

instruments (especially when trained interviewers are available) in order to improve the quality of the study, but, long before this, it needs to be asked who is providing the information and why. Thus, for example, if an intervention designed to reduce depression is being tested, it is appropriate to use a structured interview schedule of established reliability (e.g. Schedule for Affective Disorders and Schizophrenia; Endicott & Spitzer, 1978) for assessment, but if the person concerned has relatively mild symptoms that could be hidden from a stranger, it would be more appropriate to assess the patient with a self-rating scale (e.g. Hamilton Rating Scale for Depression; Hamilton, 1960).

Almost all psychiatric symptoms have both a subjective element and an objective one that is shown to others. In some instances there may be a gross disparity between the two (e.g. in the features of psychopathy), but it is rare to have one feature only. For this reason many investigators use both self-rating questionnaires and more 'objective' rating scales, although in practice these often show good levels of agreement.

One of the main advantages of the questionnaire is that it reduces the potential for bias because a patient is more likely to describe their own feelings accurately than an investigator who is involved in a comparison of treatments and has some knowledge of what these are. Often bias is unwitting and one advantage of recording both self-rated and observer-rated symptoms is that similar results with both types of instrument suggest a minimum of bias.

Devising your own instrument

Although there is a natural tendency for researchers to develop their own instruments on the premise that there is no scale available to measure a particular feature, this position is increasingly untenable as instruments become available for all aspects of psychiatric illness and treatments. There is also considerable concern that new and untested scales yield much larger effect sizes than well-established scales (i.e. overstate the difference between treatments; Marshall *et al*, 2000).

Although there are still circumstances when a new rating scale might be necessary for a specific project, it is important for researchers to be aware that such a scale should be evaluated and the results of the evaluation should be published before the scale is used in the planned study. This will invariably involve much more work than using an established scale. Nobody should believe that using a specially derived scale for a project is going to be a short cut to success.

In deciding on a new rating scale the investigator will have to make a distinction between a simple dichotomous scale, an interval scale and a visual analogue scale (Fig. 2). There is often a wish to modify an existing scale and although under some circumstances this is justifiable, it must not be done without a great deal of thought, as comparisons with data using the original scale would thereby be rendered invalid.

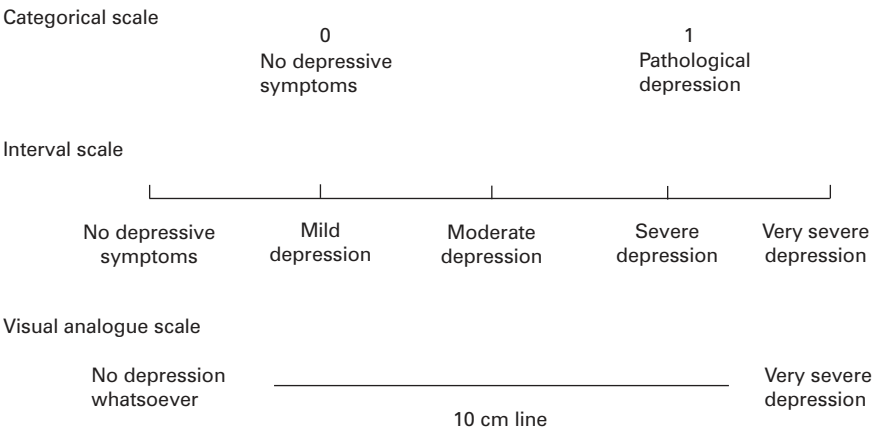

Fig. 2 Examples of types of rating scale: categorical scale, interval scale (implying dimensions, e.g. Likert scale) and visual analogue scale (the participant is asked to place a vertical mark across the line at the point that best describes current feelings; this is measured to give a 'depression score').

Finding a rating scale

The rest of this booklet lists the main scales for each area of psychiatry. This is a repetitive exercise but it is clear from talking to novice researchers that the listing of these scales is important. We decided that the main criterion for the inclusion a scale should be the extent of its use (as the wider the use of a scale the better will be comparability with other studies). We have therefore calculated the citation rate per year of each scale since the year of publication and only those scales that are widely cited in the literature (with a cut-off point of 4.0 per year for general scales and 2.0 per year for specific ones) have been included. Although we are well aware that some of the most commonly used scales are not quite as good as some others and have only achieved their status by a combination of primacy, luck and salesmanship, their frequency of use is still the best single criterion for the research worker in making a choice. Where the details of scales are not available in the published references the researcher is advised to search for these on the internet. This is now much easier with improved search engines such as Google, and any scale which is searched for using the author's name and title of the scale should be found easily. The most popular scales are frequently copyrighted and distributed by commercial publishers. For those that are less widely used but seem to be appropriate for a study it does no harm to get in touch with the originator(s). They will be flattered (unless of course the scale is so widely used it has led to many previous enquiries) and may offer extra help in starting the project. This may even be worthwhile

Table 1 Rating scales and questionnaires for depression

Author(s)	Type of assessment	Citation rate per year and comments
Hamilton (1960)	Hamilton Rating Scale for Depression (HRSD)	199.5 (the original and, to many, still the best)
Beck *et al* (1961)	Beck Depression Inventory (BDI)	186.2 (competing for the crown with enthusiasm – generally preferred in more recent studies)
Zigmond & Snaith (1983)	Hospital Anxiety and Depression Scale (HAD)	133.0 (currently the most frequently used self-rating scale, equally good for anxiety)
Montgomery & Åsberg (1979)	Montgomery–Åsberg Depression Rating Scale (MADRS)	83.2 (derived from the Comprehensive Psychopathological Rating Scale (CPRS) and may be of special value when multiple pathology is being assessed; very often used in short-term studies of interventions, particularly drugs)
Zung (1965)	Zung Self-Rating Depression Scale	78.9 (the original self-rating scale; still widely used)
Brink *et al* (1982)	Geriatric Depression Scale (GDS)	71.9 (clear preference for this scale in studies of older adults)
Beck *et al* (1974*b*)	Hopelessness Scale	38.9 (very frequently used in studies of suicide)
Cox *et al* (1987)	Edinburgh Postnatal Depression Scale (EPDS)	33.1 (the established scale for assessing depression in relationship to childbirth) Also see Cox & Holden (2003)
Seligman *et al* (1979)	Attributional Style Questionnaire	22.4
Alexopoulus *et al* (1988)	Cornell Scale for Depression in Dementia	22 (an example of a special area in which a general scale may not be accurate)
Carney *et al* (1965)	ECT Scale (Newcastle)	21.4 (was once very widely used but less so recently, as the distinction between depressive syndromes is less often required)
Kandel & Davies (1982)	Six-Item Depression Mood Inventory	15.4
Brown & Harris (1978)	Life Events and Difficulties Scale (LEDS)	11.2 (the definitive life events assessment scale – needs prior training – listed here as the work was primarily concerned with depression)
Zuckerman (1960)	Multiple Affect Adjective Checklist (MAACL)	8.5. (checklists used to be very common methods of assessing mood states but are now less often used)

continued