Likert scales: how to (ab)use them
Susan Jamieson

Dipping my toe into the water of educational research, I have recently used Likert-type rating scales to measure student views on various educational interventions. Likert scales are commonly used to measure attitude, providing ‘a range of responses to a given question or statement’. Typically, there are 5 categories of response, from (for example) 1 = strongly disagree to 5 = strongly agree, although there are arguments in favour of scales with 7 or with an even number of response categories. 

Likert scales fall within the ordinal level of measurement. That is, the response categories have a rank order, but the intervals between values cannot be presumed equal. Methodological and statistical texts are clear that for ordinal data one should employ the median or mode as the ‘measure of central tendency’ because the arithmetical manipulations required to calculate the mean (and standard deviation) are inappropriate for ordinal data, where the numbers generally represent verbal statements. In addition, ordinal data may be described using frequencies/percentages of response in each category. 

Correspondence: Dr Susan Jamieson, Section of Squamous Cell Biology and Dermatology, Division of Cancer Sciences and Molecular Pathology, Faculty of Medicine, Robertson Building, 56 Dumbarton Road, Glasgow G11 6NU, UK. Tel: 00 44 141 330 4018; Fax: 00 44 141 330 4098; E-mail: Susan.jamieson@clinmed.gla.ac.uk. doi: 10.1111/j.1365-2929.2004.02012.x

© Blackwell Publishing Ltd MEDICAL EDUCATION 2004; 38: 1212–1218
Delving further, treating ordinal scales as interval scales has long been controversial (discussed by Knapp\(^8\)) and, it would seem, remains so. Thus, while Kuzon Jr et al.\(^9\) contend that using parametric analysis for ordinal data is the first of ‘the seven deadly sins of statistical analysis’, Knapp\(^8\) sees some merit in the argument that sample size and distribution are more important than level of measurement in determining whether it is appropriate to use parametric statistics. Yet even if one accepts that it is valid to assume interval status for Likert-derived data, data sets generated with Likert-type scales often have a skewed or polarised distribution (e.g. where most students ‘agree’ or ‘strongly agree’ that a particular intervention was useful, or where students have polarised views about a ‘wet lab’ in biochemistry, depending on their interest in basic science).

It seems to me that if we want to ‘raise the quality of research’ in medical education,\(^10\) issues such as levels of measurement and appropriateness of mean, standard deviation and parametric statistics should be considered at the design stage and must be addressed by authors when they discuss their chosen methodology. Knapp\(^8\) gives advice that essentially boils down to this: the researcher should decide what level of measurement is in use (to paraphrase, if it is an interval level, for a score of 3, one should be able to answer the question ‘3 what?’); non-parametric tests should be employed if the data is clearly ordinal, and if the researcher is confident that the data can justifiably be classed as interval, attention should nevertheless be paid to the sample size and to whether the distribution is normal.

Finally, is it valid to assume that Likert scales are interval-level? I remain convinced by the argument of Kuzon Jr et al.,\(^9\) which, if I may paraphrase it, says that the average of ‘fair’ and ‘good’ is not ‘fair-and-a-half’; this is true even when one assigns integers to represent ‘fair’ and ‘good’!

references