9 Lehmann U, Blom W, Dlanjwa M et al. Capacity development through reflective practice and collaborative research among clinic supervisors in rural South Africa – a case study. *Educ Health* 2004;**17**:53–61.

10 Rogers E. *Diffusion of Innovations.* New York: The Free Press 1995.

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'.[1] 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.[1]

> *The response categories in Likert scales have a rank order, but the intervals between values cannot be presumed equal*

Likert scales fall within the ordinal level of measurement.[2–4] That is, the response categories have a rank order, but the intervals between values cannot be presumed equal, although, as Blaikie[3] points out, '...researchers frequently assume that they are'. However, Cohen et al.[1] contend that it is 'illegitimate' to infer that the intensity of feeling

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 4008; E-mail: Susan.jamieson@clinmed.gla.ac.uk.

between 'strongly disagree' and 'disagree' is equivalent to the intensity of feeling between other consecutive categories on the Likert scale. The legitimacy of assuming an interval scale for Likert-type categories is an important issue, because the appropriate descriptive and inferential statistics differ for ordinal and interval variables[1,5] and if the wrong statistical technique is used, the researcher increases the chance of coming to the wrong conclusion about the significance (or otherwise) of his research.

> *The mean (and standard deviation) are inappropriate for ordinal data*

Methodological and statistical texts are clear that for ordinal data one should employ the median or mode as the 'measure of central tendency'[5] because the arithmetical manipulations required to calculate the mean (and standard deviation) are inappropriate for ordinal data,[3,5] where the numbers generally represent verbal statements. In addition, ordinal data may be described using frequencies/percentages of response in each category.[3] Standard texts also advise that the appropriate inferential statistics for ordinal data are those employing non-parametric tests, such as chi-squared, Spearman's Rho, or the Mann–Whitney U-test[1] because

parametric tests require data of interval or ratio level.[2,5]

> *Treating ordinal scales as interval scales has long been controversial*

However, these 'rules' are commonly ignored by authors, including some who have published in *Medical Education*. For example, the authors of 2 recent papers had used Likert scales but described their data using means and standard deviations and performed parametric analyses such as ANOVA.[6,7] This is consistent with Blaikie's observation that it has become common practice to assume that Likert-type categories constitute interval-level measurement.[3] Generally, it is not made clear by authors whether they are aware that some would regard this as illegitimate; no statement is made about an assumption of interval status for Likert data, and no argument made in support.

> *It has become common practice to assume that Likert-type categories constitute interval-level measurement*

All of which is very confusing for the novice in pedagogical research. What approach should one take when specialist texts say 1 thing, yet actual practice differs?

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).

> *Such issues as levels of measurement and appropriateness of mean, standard deviation and parametric statistics should be considered in the design stage and must be addressed by authors*

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

1 Cohen L, Manion L, Morrison K. *Research Methods in Education.* 5th edn. London: RoutledgeFalmer 2000.

2 Pett MA. *Non-parametric Statistics for Health Care Research.* London: SAGE Publications 1997.

3 Blaikie N. *Analysing Quantitative Data.* London: Sage Publications 2003.

4 Hansen JP. CAN'T MISS – conquer any number task by making important statistics simple. Part 1. Types of variables, mean, median, variance and standard deviation. *J. Health Care Qual* 2003;**25** (4):19–24.

5 Clegg F. *Simple Statistics.* Cambridge: Cambridge University Press 1998.

6 Santina M, Perez J. Health professionals' sex and attitudes of health science students to health claims. *Med Educ* 2003;**37**:509–13.

7 Hren D, Lukic IK, Marusic A, Vodopivec I, Vujaklija A, Hrabak M, Marusic M. Teaching research methodology in medical schools: students' attitudes towards and knowledge about science. *Med Educ* 2004;**38**:81–6.

8 Knapp TR. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nurs Res* 1990;**39**:121–3.

9 Kuzon WM Jr, Urbanchek MG, McCabe S. The seven deadly sins of statistical analysis. *Ann Plastic Surg* 1996; **37**:265–72.

10 Bligh J. Ring the changes: some resolutions for the new year and beyond. *Med Educ* 2004;**38**:2–4.