



Jamieson, S. (2004) Likert scales: how to (ab)use them? *Medical Education*, 38(12), pp. 1217-1218. (doi: [10.1111/j.1365-2929.2004.02012.x](https://doi.org/10.1111/j.1365-2929.2004.02012.x))

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/59552/>

Deposited on: 27 April 2020

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Title: Likert Scales: How To (Ab)Use Them?

Descriptive title : Commentary on the appropriate use of data derived from
Likert-type rating scales

Word count 749 words (excluding references & pull-out quotes)

Dipping my toe in 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 five categories of response, from (for example) 1=strongly disagree to 5 =strongly agree, although there are arguments in favour of scales with seven, or with an even number of response categories¹.

Likert scales fall within the ordinal level of measurement^{2,3,4}. That is, the response categories have a rank order, but the intervals between values cannot be presumed equal, although as Blaikie³ 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 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.

Methodological and statistical texts are clear that for ordinal data one should employ the median or mode as the ‘measure of central tendency’⁵ since the arithmetic 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 category³. Standard texts also advise that the appropriate

inferential statistics for ordinal data are those employing non-parametric tests, such as Chi-square, Spearman's Rho, or the Mann-Whitney U-test¹, since parametric tests require data of interval or ratio level^{2,5}.

However, these 'rules' are commonly ignored by authors, including some who have published in Medical Education. For example, in two recent papers the authors have used Likert scales, but have described their data using means and standard deviations and have 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 measurement³. 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.

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

Delving further, treating ordinal scales as interval scales has long been controversial (discussed by Knapp⁸) 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", Knapp⁸ 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 education¹⁰, 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 when they discuss their chosen methodology.

Knapp⁸ gives advice that essentially boils down to this: the researcher should decide what level of measurement is in use (to paraphrase, if it’s interval level, for a score of 3, one should be able to answer “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: to paraphrase, the average of ‘fair’ and ‘good’ is not ‘fair-and-a-half’, and 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. *Nonparametric Statistics for Health Care Research*. London: SAGE Publications, 1997.
3. Blaikie N. *Analyzing Quantitative Data* London: SAGE Publications Ltd., London, 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. Healthcare Qual* 2003; **25** (4): 19-24.
5. Clegg F. "Simple Statistics", Cambridge University Press, 1998.
6. Santana M, Perez J. "Health professionals' sex and attitudes of health science students to health claims." *Med Educ*. 2003; **37**: 509-513.
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**: 81-86.
8. Knapp, TR "Treating ordinal scales as interval scales: an attempt to resolve the controversy" *Nursing Res*. 1990; **39**: 121-123.
9. Kuzon, Jr. WM, Urbanek MG, McCabe S. "The seven deadly sins of statistical analysis" *Annals Plastic Surg*. 1996; **37**: 265-272.
10. Bligh J "Ring the changes: some resolutions for the new year and beyond" *Med. Educ*. 2004; **38**: 2-4.

POTENTIAL 'PULL-OUT' QUOTATIONS

“Likert scales fall within the ordinal level of measurement ... the response categories have a rank order, but the intervals between values cannot be presumed equal”

“... the mean (and standard deviation) are inappropriate for ordinal data”

“Treating ordinal scales as interval scales has long been controversial”

“...it has become common practice to assume that Likert-type categories constitute interval-level measurement”

“...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...”