Resolving the 50-year debate around using and misusing Likert scales

James Carifio & Rocco Perla

How Likert type measurement scales should be appropriately used and analysed has been debated for over 50 years, often to the great confusion of students, practitioners, allied health researchers and educators. Basically, there are two major competing views that have evolved somewhat independently of one another and of the associated empirical research literature on this 'great debate'. Most recently in this journal, Jamieson[1] outlined the view that 'Likert scales' are *ordinal* in character (i.e., produce rank order data) and that they, therefore, must be analysed using non-parametric statistics. Non-parametric statistics, however, are less sensitive and less powerful than parametric statistics and are, therefore, more likely to miss weaker or emerging findings.

> *Historically, there has been debate between those who maintain the ordinalist (rank order) and intervalist views in Likert scales*

Pell[2] responded to Jamieson's article with a letter that pointed out three of the chief reasons why Likert *scales* (collections of items) as opposed to individual Likert *items* are not ordinal in character, but rather

Lowell, Massachusetts, USA

Correspondence: James Carifio, University of Massachusetts-Lowell, Lowell, Massachusetts, 01854, USA. Tel: 00 1 617-513-6279; Fax: 00 1 978-934-3005;
E-mail: james_carifio@hotmail.com

are *interval* in nature and, thus, may be analysed parametrically with all the associated benefits and power of these higher levels of analyses. Jamieson[3] replied that she failed to be convinced by Pell's three points. Although both views are prototypical of this long and ongoing debate, neither the letter nor the original commentary made use of a great deal of empirical evidence that should enable a resolution.

> *Monte Carlo studies of the F-test have shown the F-test to be extremely robust to violations of its assumptions, which must be extreme before the F-test is biased*

Monte Carlo studies of the *F*-test, performed by Glass *et al.*,[4] have convincingly shown that the *F*-test is extremely robust to violations of its assumptions, except for the homogeneity of variance assumption, and violations of this assumption must truly be extreme before they bias the *F*-test. Utilising the *F*-test to analyse ordinal data, therefore, produces unbiased results, which is an *empirical fact*. Further, a variety of studies on the nature of Likert *scales* (as opposed to single Likert items) have shown that the Likert response format produces *empirically interval data*[5,6] and, in fact, can approximate *ratio* data, in theory and actuality , if a hundred millimeter response line is used for marking responses which has 'always' and 'never' as anchors.[7] The weight of the empirical evidence, therefore, clearly supports the view and position that Likert *scales* (collections of Likert items)

produce *interval data*, particularly if the scale meets the standard psychometric rule-of-thumb criterion of comprising at least eight reasonably related items. Based on the empirical evidence available, then, the argument clearly goes to Pell and the *intervalist* view of the issues he represents in this debate. However, to resolve the debate we must also ask where this contention that Likert scales produce ordinal data, and all that this unsupported contention implies, comes from, and why it has persisted as a view long past the point when adequate data and arguments became available to decide the debate.

> *A variety of studies have shown that the Likert response format produces empirically interval data at the scale level*

The views of Likert scales and their uses and misuses presented by Jamieson[1] can be characterised as representative of Stevens' non-parametric view (and fallacy). Stevens'[8,9] view is that what is or might be ordinal data at the item (i.e. atom) level cannot be interval data at the scale (i.e. molecular) level, whereas all in medicine know that molecules almost always have properties that their individual atoms do not have but are functionally reliant on them nonetheless. It is the 'emergent properties' of scales (versus items) that are key in framing and settling disagreements in this area.

Stevens' view is a logical argument based upon extrapolating various

rank ordering methodologies he was investigating at the time to 'Likert scales', which, parenthetically, he and others mischaracterised in several ways relative to what Likert wrote and data that were available at the time.[10] Empirical data and studies often do not significantly impact logical arguments or logical theories, even if we acknowledge that the best empirical evidence did not become available until 20 years after this view had become firmly rooted in several disciplines and communities.

> *Those who hold the ordinalist view of Likert scales rarely mention the abundant empirical findings about Likert scales*

Those who hold the ordinalist (i.e., Stevens') view of Likert scales, and of how the data from these scales should be analysed, rarely mention or address the empirical findings and facts outlined in this commentary. These facts and findings are simply ignored, or seem not to be familiar to those holding the ordinalist view of Likert scales. Beliefs that live on independently of empirical evidence and key studies are typically called misconceptions, myths or urban legends. We have identified and codified 10 major misconceptions, myths and urban legends about Likert scales and their analysis in detail elsewhere, and include in that paper a wide array of additional supporting empirical evidence.[11]

In the article cited above[11], we focused on a number of the chief misconceptions and logical flaws in the ordinalist position in this 50-year debate about Likert scales because these misconceptions and logical flaws are as important as the empirical evidence that strongly falsifies this view. For example, the ordinalist view makes no distinction between a Likert *response format*, a

Likert (graded valence) *question* (or stem) or a Likert *scale* (collection of items), which are the root of many of the logical problems with the ordinalist position and many of the incorrect claims ordinalists make in this debate. The intervalist position, by contrast, brings all of these (and other) elements together to form a true and coherent measurement system whose construct validity may be easily and quickly assessed, representing another critical difference between the two views.

We believe that reading the article we have written on this topic and the sources it cites will convince most paradigmatically neutral researchers and practitioners of the correctness of the points made and supported by the article, and the correctness of the view of the debate we summarise in this commentary. Likert methodology is one of the most commonly used methodologies in all fields of research, but particularly so in allied health, medicine and medical education.

> *It is perfectly appropriate, therefore, to sum Likert items and analyse the summations parametrically, both univariately and multivariately*

Therefore, we do not want to see researchers and practitioners unwittingly misusing and misunderstanding Likert scales, their nature and characteristics, or how data obtained using them should be analysed with maximal sensitivity and power. The debate on Likert scales and how they should be analysed, therefore, clearly and strongly goes to the *intervalist position*, if one is analysing more than a *single* Likert item. Analysing a *single* Likert item, it should also be noted, is a practice that should only occur very rarely. It is, therefore, as the

intervalists contend, perfectly appropriate to summarise the ratings generated from Likert *scales* using means and standard deviations, and it is perfectly appropriate to use parametric techniques like Analysis of Variance to analyse Likert scales. It is also perfectly appropriate to calculate Pearson correlation coefficients using the summative ratings from Likert scales and to use these correlations as the basis for various multivariate analytical techniques, such as multiple regression, factor analysis and meta-analysis, to obtain more powerful and nuanced analyses of the data and research hypotheses being investigated. Treating the data from Likert *scales* as ordinal in character prevents one from using these more sophisticated and powerful modes of analyses and, as a result, from benefiting from the richer, more powerful and more nuanced understanding they produce.

REFERENCES

1 Jamieson S. Likert scales: how to (ab)use them. *Med Educ* 2004;**38**:1212–8.
2 Pell G. Uses and misuses of Likert scales. *Med Educ* 2005;**39**:97.
3 Jamieson S. Author's reply. *Med Educ* 2005;**39**:970.
4 Glass GV, Peckham PD, Sanders JR. Consequences of failure to meet assumptions underlying the analyses of variance and covariance. *Rev Educ Res* 1972;**42**:237–88.
5 Carifio J. Assigning students to career education programs by preference: scaling preference data for program assignments. *Career Education Quarterly* 1976;**1,1**, Spring, 7–26.
6 Carifio J. Measuring vocational preferences: ranking versus categorical rating procedures. *Career Education Quarterly* 1978;**3,1**, Winter, 34–66.
7 Vickers A. Comparison of an ordinal and a continuous outcome measure of muscle soreness. *Int J Technol Assess Health Care* 1999;**15**:709–16.

8 Stevens S. On the theory of scales of measurement. *Science* 1946;**103** (67):668–90.

9 Stevens S. Mathematics, measurement and psychoanalysis. In: Stevens SS, ed. *Handbook of Experimental Psychology.* New York: John Wiley & Sons 1951;1–49.

10 Likert R, Hayes S. *Some Applications of Behavioural Research.* Paris: Unesco 1957.

11 Carifio J, Perla R. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of the Social Sciences* 2007;**3**(3), 106–116.

Thinking 'no' but saying 'yes' to student presence in general practice consultations: politeness theory insights

Charlotte E Rees[1] & Lynn V Knight[2]

We've all done it. We've all agreed to something when we really wanted to say no. It was no surprise to us, therefore, that 20.8% of patient participants who agreed to have a student present during their consultation in the Price *et al.*[1] study would have preferred to see their general practitioner (GP) alone. This finding is consistent with results from our qualitative research in which medical students articulated their anxieties about patient consent.[2] Although students were keen for patients to give consent because this would provide the students with learning opportunities, they were worried that patients sometimes gave consent without understanding what they were consenting to and without really wanting to consent. This tension between patient consent and student learning has been articulated in earlier literature[3] and is illustrated in turns 1 and 5 of the previously unpublished excerpt from Knight and Rees.[2]

> *The patient's ability to refuse a request for student presence relates to who asks for consent (and the patient's relationship with the requester) and who listens to the response*

This excerpt raises two important issues, the resolution of which may help us understand the conditions necessary for some patients to agree to the presence of medical students despite their wish to say no. These issues concern who requests consent (and the patient's relationship with the requester) and who hears the patient's response. From Ann's comments in turn 5, we can see that patients seem able to decline a request for students to be present if they are asked by a receptionist (a third party extrinsic to the patient–doctor relationship) and if their response will not be heard by either the GP or the student. By contrast, Matt implies that patients are unable to decline a request if they are asked by the GP (a person central to the patient–doctor relationship) and if their response will be heard by the student. So, why is it that patients feel unable to decline a GP's request when the student is present to hear the response? To answer this question, we turn to politeness theories.[4,5]

Drawing on Goffman,[6] Brown and Levinson[4] employed the concept of *face* (a person's public self-image) to illustrate how people go to great efforts in social interaction to maintain their own and others' self-image needs using politeness strategies. They differentiate between positive face (maintaining a positive self-image) and negative face (maintaining autonomy as in freedom from imposition and freedom of choice) and suggested that speech acts, such as requests and refusals, threaten interlocutors' positive and negative face needs. Requests not only threaten the requestee's freedom from imposition,[4] but threaten multiple faces, including the requester and requestee's positive self-image.[5] Likewise, refusals threaten the requester's autonomy and other types of face, depending on the reasons behind the refusal (known as obstacles).[7]

> *Requests and refusals not only threaten the negative face of the recipient but threaten multiple faces*

[1]Sydney, New South Wales, Australia
[2]Cardiff University, Cardiff, UK

Correspondence: Charlotte E Rees, OPME, Mackie Building (K01), University of Sydney, Sydney, New South Wales 2006, Australia. Tel: 00 61 (0) 2 9351 2814; Fax: 00 61 (0) 2 9351 6646; E-mail: crees@med.usyd.edu.au