Peter Grimbeek
Griffith University

Book Review

Thirteen good reasons for applying the Rasch model to human data

Bond, Trevor, G. & Fox, Christine M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

ISBN hardback: 0-8058-3476-1; ISBN paperback 0-8058-4252-7
Price: Australian Price: $59.35 (AUD) $65.28 (AUD) including GST
Distribution: DADirect; www.dadirect.com.au; Phone: 61 3 9210 7777; Fax: 61 3 9210 7788
Book website: http://www.jcu.edu.au/~edtgb

Michell (1999) wrote eloquently about problems related to measurement properties commonly ignored in social science research (Grimbeek, 1999). Bond and Fox acknowledge these concerns and outline solutions based on the Rasch model, a model based on item response theory. They provide examples of the steps in analysis, the outputs, and the control files for various Rasch analytic procedures. Despite its avowedly technical nature, the clarity of the exposition recommends this book to the moderately numerate reader (i.e., most of us). Researchers may not agree that Rasch modelling is the only solution but will come away with additional options for research and analysis.

In the first two chapters, Bond and Fox ask questions about the level of measurement usually available in the social sciences as a preface to introducing the Rasch model. In Chapter 1, they remind us that Stevens' (1946) definition of measurement as "the assignment of numerals to objects or events according to a rule" is problematic and that the ideal of equal interval measurement scales for the most part remains just that. In Chapter 2, they outline the way in which the Rasch model delivers measurement scales with equal interval properties. The secret, it seems, is to order the "dummy" coded dichotomous data set in terms of least to most occurrences (for persons and items), to calculate item and person abilities as fractional scores (representing success-to-failure ratios), and then to transform the raw score summaries to natural logarithms. An important claim is made that the resulting distribution of scores (and persons) has the properties of an (equal) interval scale.

https://doi.org/10.1017/S0816512200028315 Published online by Cambridge University Press

In the remainder of the book, they discuss applications of the Rasch model. Early chapters address issues relating to the analysis of (a) dichotomous responses, (b) Likert scale items, and (c) items with differing numbers of response categories. Later chapters examine analyses that take into account (d) differences between raters, (e) longitudinal data, and (f) model fitting considerations.

One important issue is whether a set of items represents a single factor. As stated early in Chapter 3, the Rasch model assumes a developmental pathway analogy. That is, the Rasch model requires (the practitioner assumes) that items test a single underlying factor, so that differences in the frequency of "correct" answers reflect differences in item difficulty or personal ability rather than the presence of differing factors. That is, it is presumed that, rather than deriving from differing factors, communalities and differences derive from clusters of items at particular difficulty levels or clusters of test takers at particular ability levels. Because many measurement instruments either are explicitly multidimensional or include dubious items, one might be excused for wanting to combine Rasch analytic procedures with factor analytic procedures that survey the possibility of multidimensional factor structures. Rasch analysis does allow for the possibility of nonunitary factor structures to the extent that in-fit and out-fit statistics can be interpreted as indicating that not all items measure a given (unidimensional) developmental pathway equally well. However, if items that form part of a particular instrument are carefully selected for face and construct validity with the intention of detecting various amounts of an underlying dimension, then the initial requirement for unidimensionality is more likely to be met in practice.

In Chapters 6-9, Bond and Fox show how the Rasch model, applied in earlier chapters to dichotomous variables, generalizes to more complex rating schemes, most notably Likert scale items. They make a strong case for applying Rasch analytic procedures to Likert scale items. Research using Likert scale questionnaires (Likert, 1932) for attitude testing, etc, is common in social science research and, as such, constitutes an excellent example of how research methods can affect outcomes. The "rule of thumb" seems to be that a continuum of response categories such as "strongly disagree", "disagree", "undecided", "agree", and "strongly agree" can be recoded numerically as "1", "2", "3", "4", and "5." The resulting set of numbers is then commonly analysed as if possessing equal interval properties. Participant responses signifying, for example, agreement versus disagreement typically are analysed by computing central tendency and range (descriptive statistics) and by examining significant differences between conditions (inferential statistics).

The Rasch model's approach to examining Likert scale items differs from this typical analysis in that it takes explicit account of the distinction between ordinal and interval measurement scales. In addition, Bond and Fox contribute the useful insight that individual responses to an item on, for

https://doi.org/10.1017/S0816512200028315 Published online by Cambridge University Press

example, a 5-step response scale representing increasing level of agreement can usefully be analysed as five sets of responses. That is, each step in level of agreement can be treated as a separate dichotomous variable. For instance, one could count the number of *strongly disagree* responses for a particular item as one dichotomous response to that item. In principle, five questionnaire items that share a common 5-step Likert response scale yield 25 separate data points. This multiplicative increase in the number of data points also requires a correspondingly large sample of respondents but has the virtue of giving a suitably subtle answer to the complex question we pose when we collect information in the form of Likert scale items.

Bond and Fox demonstrate the ability of Rasch analysis to answer questions posed by Likert scale items in examining student responses to the Computer Anxiety Survey (CAS) developed by Simonson *et al.* (1987). The CAS consists of 26 Likert-type items considered to be indicative of a person's feelings of anxiety towards computers. King (1993) used conventional statistical procedures to obtain counterintuitive evidence that 120 Year 7 student CAS scores actually increased over a 9-month period of exposure to computers (i.e., one would expect anxiety to decrease with increasing experience of computers). Bond and King (1996) show the dangerous simplicity of this approach in the course of a Rasch analysis of the responses of 327 Year 7 students to this computer anxiety index. Rasch analysis demonstrated that not all items appear to be revealing the same construct, that some items exhibit a lower threshold in terms of computer anxiety than others, and that many of the test items appear to be too "difficult" in the sense that most test takers did not appear at all anxious. One outcome of this Rasch based analysis of the Computer Anxiety Survey is to provide reasons to doubt the validity of the initial report of an increase in computer anxiety.

One point of particular interest to me, in the concluding chapters, is the suggestion that Rasch procedures can be used to provide guidelines for collapsing the responses to Likert scale items across response categories (e.g., collapsing across *agree* and *strongly agree*). Historically speaking, Likert scale items (Likert, 1932) emerge naturally from the use of dichotomous items. Responses to dichotomous items, when restated as dummy variables (0,1 values), are, strictly speaking, linear (Tabachnick & Fidell, 1996) and thus, in principle, able to be analysed with parametric procedures. (Note, however, that Rasch analysts view dichotomous items as requiring further rescaling prior to being ready for parametric analysis). Yet responses to such dichotomous items are likely to exhibit skewed distributions because of the truncated response range. Skewed data sets present problems to analysts in that parametric procedures also assume normalcy of distribution. In this regard, Likert response scales that spread "*agree* versus *disagree*" responses across 4- to 6-point scales (cf., Babbie, 1995) presented a way to lessen the frequency of skewed responses. In sum, Likert scales can be viewed as a clever attempt to transform skewed dichotomous scores into more normally distributed scores.

Viewed in this light, the further step of collapsing multiple Likert scale response categories into, say, trichotomous or dichotomous categories is in keeping with this book's assertion that the multiple responses generated by Likert scale items create problems for analysis that are best resolved by either transforming the existing range of responses to produce interval scale responses or by reducing the number of response categories or both. The practice of collapsing response categories, which has been around since at least the 1950s (e.g., Selvin, 1970), is one eminently sensible way out of the impasse of a multiplicity of response categories. Setting aside the thought of using Rasch procedures to make these decisions, the rule of thumb for so doing is precisely as stated by Bond and Fox, that the emergent categories must make sense and must have enough observations in each category.

The Rasch model may not be the only way to rescale item scores so as to form scales with interval properties (e.g., factor analytic procedures can be used to transform item scores into interval-scale factor scores). However, the model does offer definite advantages, many outlined within the pages of this splendidly readable 13-chapter book.

References

Babbie, E. (1995). *The practice of social research* (7th ed.). Place, USA: Wadsworth.

Blampied, N. (1999). A legacy neglected: Restating the case for single case research in cognitive-behaviour therapy. *Behaviour Change, 16*, 89-104.

Grimbeek, P. (1999). Reasons for reconsidering quantitative research based on the use of Likert Scale and other Social Science data sets. *The Australian Educational and Developmental Psychologist, 16*(2), 89-91.

King, J., & Bond, T. (1996). A Rasch analysis of a measure of computer anxiety. *Journal of Educational Computing Research, 14*(1), 49-65.

King, J. (1993). Getting anxious about electronic learning. *Australian Educational Computing, 8*(2), 16-20.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1-54.

Michell, J. (1999). *Measurement in psychology: Critical history of a methodological concept*. New York: Cambridge University Press.

Selvin, H. C. (1970). A critique of tests of significance in survey research. In D. E. Morrison & R. E. Henkel (Eds.), *The significance test controversy* (pp. 94-106). Chicago: Aldine Publishing Company.

Simonson, M. R., Maurer, M., Montag-Torardi, M., & Whitaker, M. (1987). Development of a standardized test of computer literacy and a computer anxiety index. *Journal of Educational Computing Research, 3*(2), 231-247.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 667-80.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). Place, USA: Harper Collins College Publishers.