Rasch – a look under the carpet

CAMBRIDGE ASSESSMENT

Tom Bramley

10th UK Rasch users' group, Durham

18 March 2016

UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

It makes you think…

CAMBRIDGE ASSESSMENT

"In my research I used a probabilistic model to develop an instrument to measure a latent trait."

Probabilistic

CAMBRIDGE ASSESSMENT

What does the probability in a Rasch/IRT model mean (Holland 1990)?

- Single person, single item, proportion (x=1) in hypothetical replications?
- Proportion (x=1) in group of persons at same trait level, single item?
- Proportion (x=1) in group of items at same trait level, single person?
- Your belief about p(x=1)?

Are you measuring properties of individuals, or of aggregates?

- Measurement is about individuals (Fisher, 2010)
- Trait models say nothing about individuals (Lamiell, 2013)
- If you can calibrate an instrument based on theory, you can measure individuals (Stenner et al, 2009)

Model

CAMBRIDGE ASSESSMENT

An overused word these days?

- Scientific model, statistical model, measurement model, analogical model, task model, student model, evidence model, cognitive model…

In physics (it seems) you have measurement, not measurement models

- Search for ‘measurement model’ and you get SEM – relationship between latent variable and its indicators
- Sleight-of-hand switch of one sense of ‘model’ for another? Freedman (1985)

Measurement

CAMBRIDGE ASSESSMENT

What is measurement? (Michell, 1997)

- The Rasch paradox (Michell, 2008)
- Psychological measurement is not physical measurement (Thurstone, 1927)
- Is “is X measurable?” a conceptual question (about meaning) or an empirical question (contingent on how the world is)? (Maraun,1998)

In what sense is using the Rasch model equivalent to measuring?

- Sample-free calibration (Rasch, 1961)
- Model comes first. Andrich (2004), Wright (1977), Guttman (1977)
- Units in social science measurement. Lexiles? Micromorts?

Latent trait

CAMBRIDGE ASSESSMENT

Lots of words:

- latent trait, latent variable, factor, construct, attribute, characteristic, dimension…

What is a latent trait?

- Do attributes ‘exist’? Can they ‘cause’ their indicators? (Borsboom et al, 2004)
- Category error? (Maraun & Halpin, 2008)
- Do Rasch models say anything about causation? No. (Stenner et al 2008, 2009)
- Common property of an infinite set of items? (McDonald, 2003)

Rasch and factor analysis

CAMBRIDGE ASSESSMENT

Modern (statistical) approaches subsume into a taxonomy of models

- Manifest/latent; discrete/continuous; formative/reflexive; logit/probit

Still seems to be some confusion (well, I'm confused!)

- Factor analysis of raw scores to check for unidimensionality *before* doing a Rasch analysis?! (e.g. van der Lans et al. 2015)
- Is the 'factor score indeterminacy' problem a problem for Rasch? (McDonald, 2003)
- Is it something to do with how person parameters are estimated? (Bartholomew et al., 2011)

Estimation & models

Estimation method

- Is Conditional Maximum Likelihood really that much better than Joint ML (any studies showing a substantive difference in research conclusions?)

How should you simulate scores from two examiners marking the same set of questions?

- 3-facet model conceptually flawed (though seems to work OK in practice)
- Need to distinguish 2 traits – ability of examinee and quality of response?

Disordered thresholds

CAMBRIDGE ASSESSMENT

Disordered thresholds in the Partial Credit Model

- Ongoing controversy (for the latest round see Adams et al. 2013; Andrich 2013)
- Interpretation of probability is critical. Javelin example: repeated throws of a javelin, or repeated/combined judgments of where a single throw has landed?

Measure one thing at a time

See a lot of “multidimensional Rasch model” these days…an oxymoron? (Wright & Linacre 1989)

- Fallacy to think than one trait can give information about another (at the individual level)?

Rating scale model

Should be more widely used for questionnaires with the same set of response categories for each item?

- Reasonable to think that people interpret the categories differently…
- … but is it reasonable to think that all people interpret the response categories in the same way but differently for each item (which PCM would suggest)?

Justifying the effort

CAMBRIDGE ASSESSMENT

When to use Rasch approach?

- If the scale transcends the items
- Measurement of individuals is the main aim
- Principled rules for constructing items to meet the definition of the concept (construct)
- Large ‘universe’ of items and a ‘long’ continuum
- When it gives a different answer to a plausible alternative approach, and the different answer is justifiably better

When to try something else?

- One-off instrument
- Multi-faceted concept of interest
- When data reduction / visualisation is the main aim

Final challenges

If your research involves using a probabilistic model to develop an instrument to measure a latent trait, can you say…

1. What you mean by ‘probability’, ‘model’, ‘measure’, and ‘latent trait’?

2. What rules or principles define the universe of content for the items?

3. How the instrument works?

Good luck! And thank you for listening.

bramley.t@cambridgeassessment.org.uk

www.cambridgeassessment.org.uk/our-research

References

Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement, 72*(4), 547-573.

Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care, 42*(1; Supplement: I-7).

Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any “threshold disorder controversy”. *Educational and Psychological Measurement, 73*(1), 78-124.

Bartholomew, D., Knott, M. & Moustaki, I. (2011). Latent variable models and factor analysis: a unified approach. Ch1, p16-17.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203-219.

Fisher, W. P. (2010). Statistics and measurement: clarifying the differences. *Rasch Measurement Transactions, 23*(4), 1229-1230.

Guttman, L. (1977). What is not what in statistics. *The Statistician, 26*(2), 81-107.

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika, 55*(4), 577-601.

Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology, 31*(1), 65-71.

Maraun, M. D. (1998). Measurement as a normative practice: implications of Wittgenstein's philosophy for measurement in psychology. *Theory & Psychology, 8*(4), 435-461.

Maraun, M. D., & Halpin, P. F. (2008). Manifest and Latent Variates. *Measurement: Interdisciplinary Research and Perspectives, 6*(1-2), 113-117.

McDonald, R. P. (2003). Behavior domains in theory and in practice. *Alberta Journal of Educational Research, 49*(3), 212-230.

Michell, J. (1997). Reply to Kline, Laming, Lovie, Luce and Morgan. *British Journal of Psychology 88*, 401-406.

Michell, J. (2008). Conjoint measurement and the Rasch paradox. *Theory & Psychology, 18*(1), 119-124.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (pp. 321-333). Berkeley: University of California Press.

Stenner, A. J., Stone, M. H., & Burdick, D. S. (2009). The concept of a measurement mechanism. *Rasch Measurement Transactions, 23*(2), 1204-1206.

Stenner, A. J., Stone, M. H., & Burdick, D. S. (2008, 2009). *Rasch Measurement Transactions, 22*(1), 1152-1153 and *22*(4), 1176-1177.

Thurstone, L. L. (1927). A mental unit of measurement. *Psychological Review, 34*(6), 415-423.

van der Lans, R. M., van de Grift, W. J. C. M., & van Veen, K. (2015). Developing a teacher evaluation instrument to provide formative feedback using student ratings of teaching acts. *Educational Measurement: Issues and Practice, 34(3)*, 18-27.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14(2),* 97-116.

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil, 70(12),* 857-860.