September, 2015

Historicizing Scale Development: Shifting Epistemic Practices in Rasch Measurement

Jacob Pearce, *ACER*

HISTORICIZING SCALE DEVELOPMENT:

SHIFTING EPISTEMIC PRACTICES IN RASCH MEASUREMENT

Jacob Pearce
Research Fellow
Assessment & Psychometric Research
Jacob.pearce@acer.edu.au

Overview

- The approach: the spirit of historical epistemology
- Brief history of scale development
- Rasch analysis
- Shifting epistemic practices
- Concluding remarks

The Approach: The spirit of historical epistemology

- An intellectual history of measurement, tradition of historical epistemology.
- Hans-Jörg Rheinberger—an investigation into "the historical conditions *under* which, and the means *with* which, things are made into objects of knowledge".
- Tracing the conceptual and practical conditions under which Rasch analysis, as we know it today, become possible.
- Arnold Davision—"Unless we examine the historical conditions under which our concepts emerge, we are liable to find ourselves surrounded by philosophical perplexity".
- Looking for shifting practices and changes in conceptual pre-conditions in the initial emergence and later consolidation of Rasch measurement as a powerful and ubiquitous tool in scale development and construction.

Brief History of Scale Development

- The scaling tradition (based on Item-Response Theory (IRT)) is distinct from the more traditional test-score tradition (also known as classical test theory (CTT)).
- Scaling involves the calibration of individual items located on a scale and the measurement of individual persons on the same scale against an assumed shared latent trait.
- Educational measurement—a long history of using instruments to discriminate candidates, ranking their achievement against a construct (or multiple constructs).
- In high-stakes aptitude tests, ranking candidates by locating their ability on a scale is standard practice.

Brief History of Scale Development

- Scales are also used to measure a range of latent variables; objective measures of physical aspects, subjective symptoms, functional performance, and feelings of satisfaction in surveys.
- Measures of opinions, preferences and behaviours are operationalized through measurable variables, which relate to conceptual variables.
- Measurement instruments are comprised of a collection of items which can be combined into a composite score. This score when reviewed against a scale is supposed to reveal levels of variables which are not readily observable by direct means.
- E.g., psychologists postulate constructed notions of anxiety and depression to explain observed behaviour. Measuring anxiety or depression, however, requires operationalization.

Rasch analysis

- Measurement models developed by Georg Rasch in 1960 have become widely used in educational measurement.
- However, Rasch analysis is becoming far more common as a technique in the health sciences, clinical medicine, psychology, social science, marketing and business contexts.
- The Rasch model is a powerful measurement technique for analyzing response options on collections of items, by linking candidate ability with item difficulty.
- Individuals can be compared without requiring the same items to be administered, as it is assumed that individuals and items can be located against a unidimensional latent trait.
- Rasch measures are thus generalizable across various samples and test items.

Rasch analysis

- For dichotomous data, the Rasch model takes the form:

$$Pr\{X_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

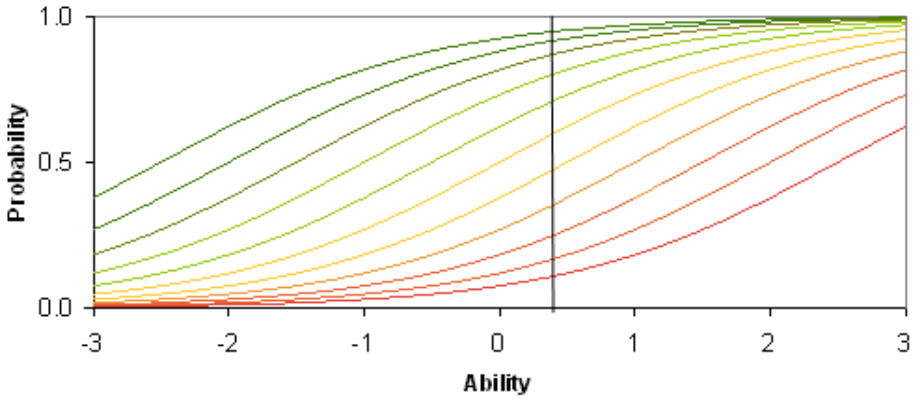
where βn is the ability of person *n*, and δi is the difficulty of item *i*.

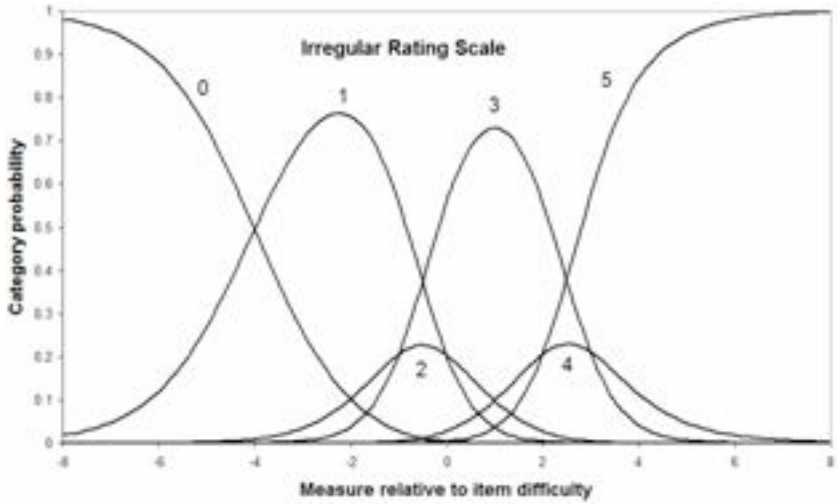
- For polytomous data, the 'partial credit' Rasch model which deals with ordered categories used in ratings scales takes the generalized form:

$$Pr\{X_{ni} = x\} = \frac{e^{-\tau_{1i} - \tau_{2i} \ldots - \tau_{xi} + x(\beta_n - \delta_i)}}{\sum_{x'=0}^{m_i} e^{-\tau_{1i} - \tau_{2i} \ldots - \tau_{x'i} + x'(\beta_n - \delta_i)}}$$

where βn is the ability of person *n*, and δi is the difficulty of item *i*, and τ_{xi}; $x = 1, 2, \ldots m_i$ are thresholds which separate the latent trait of item *i* into $m_i + 1$ distinct ordered categories.

Rasch analysis





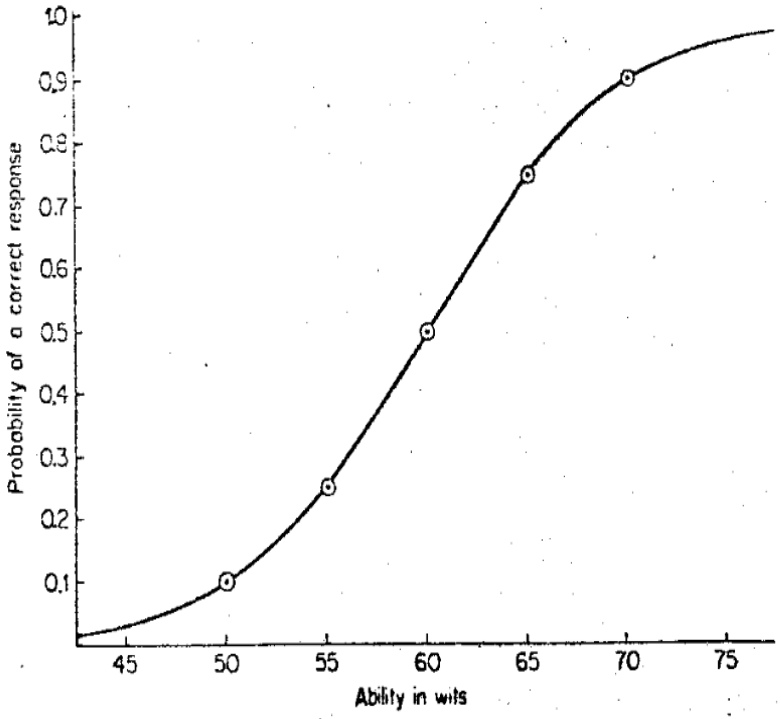
Shifting epistemic practices

New methods: shifts in educational measurement

- 1920s, educational measurement was keen on the idea of objective testing. Several issues of the *Journal of Educational Measurement* were devoted to a 1921 symposium on the scientific measurement of intelligence.
- In 1951, Frederic Lord's doctorial dissertation explored the application of latent trait theory to test theory. However, he believed that the theory would be difficult to use in practice due to the large data sets required.
- Today, using large data sets is commonplace. IRT is deployed in countless large-scale educational assessment contexts. International comparative studies such as the Programme for International Student Assessment (PISA), the Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) all use Rasch analyses.

Shifting epistemic practices

- Since emerging, Rasch measurement was constrained by the tools available to the community of practitioners.
- In order to effectively use Rasch analysis, highly competent mathematicians were required in order to manually calculate the particular Rasch estimates for items and persons.
- Plotting the results of these data was also a slow and painstaking exercise.



Shifting epistemic practices

New concepts: shifts in conceptual pre-conditions

- The 1970s (and into the 1980s) saw a plethora of research papers and new ways of applying IRT and the Rasch model.
- Measurement and psychometrically aligned journals published mathematical models—new researchers coming into the field were exposed to a variety of methods and techniques.
- 1974, Lord developed a computer program for carrying out parameter estimation with the logistic test model.
- 1977, *Journal of Educational Measurement* ran a special issue on IRT applications.
- 1979, Wright and Stone published *Best Test Design*, which described the theory underlying the Rasch model and its potential applications.
- 1979 conference on "Computerized Adaptive Testing".

Shifting epistemic practices

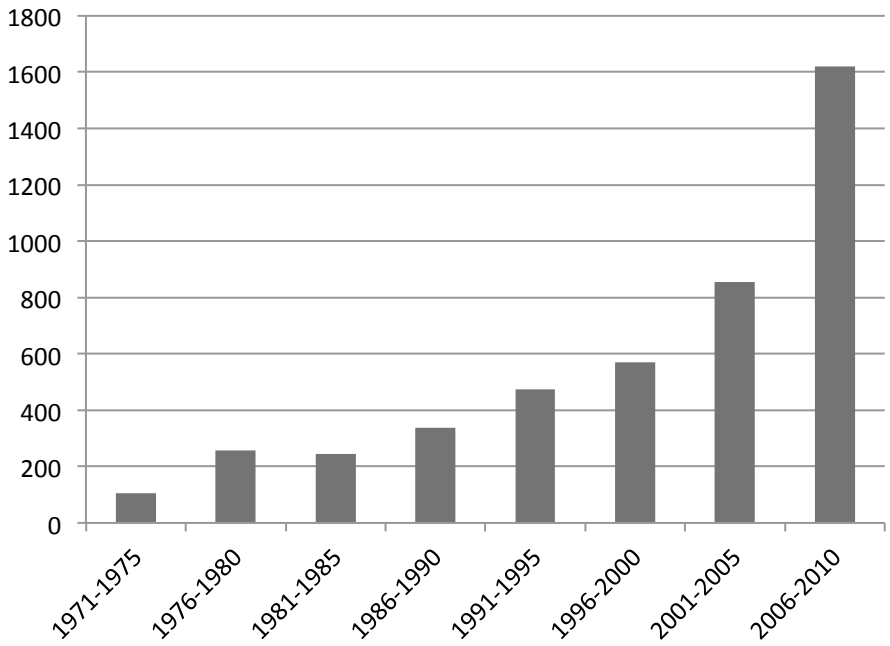
New techniques: shifts in the domain of applicability

- SPSS software in 1968, first manual in 1970, researchers given the power to manage data and undertake their own statistical analyses.
- In the 1980s advances in computing resulted in the release of sophisticated computer software to run Rasch analysis.
- 1983 Wright and Linacre released *Microscale* in MS-DOS. Followed by *Bigscale* in 1989 and later *BigSteps* in 1991 (later *Winsteps* in 1998).
- ACER's published *Quest* in 1989 and later released a broader statistical program with Rasch analysis tools built in, called *ConQuest*.
- Mid 2000s, more 'off the shelf' software was available for purchase. Programs such as *RUMM2020* were designed to be particularly user friendly for people with little mathematical training. A basic understanding of statistics was sufficient to navigate through the interface.

Shifting epistemic practices

Rasch epistemic practices today

- These new computing techniques were the defining historical pre-condition for the proliferation of Rasch analysis through the social sciences, in non-achievement contexts.
- Rasch analysis has quickly become more popular when publishing results based on scales in the social sciences.



Number of Rasch articles published by year (via GoogleScholar).

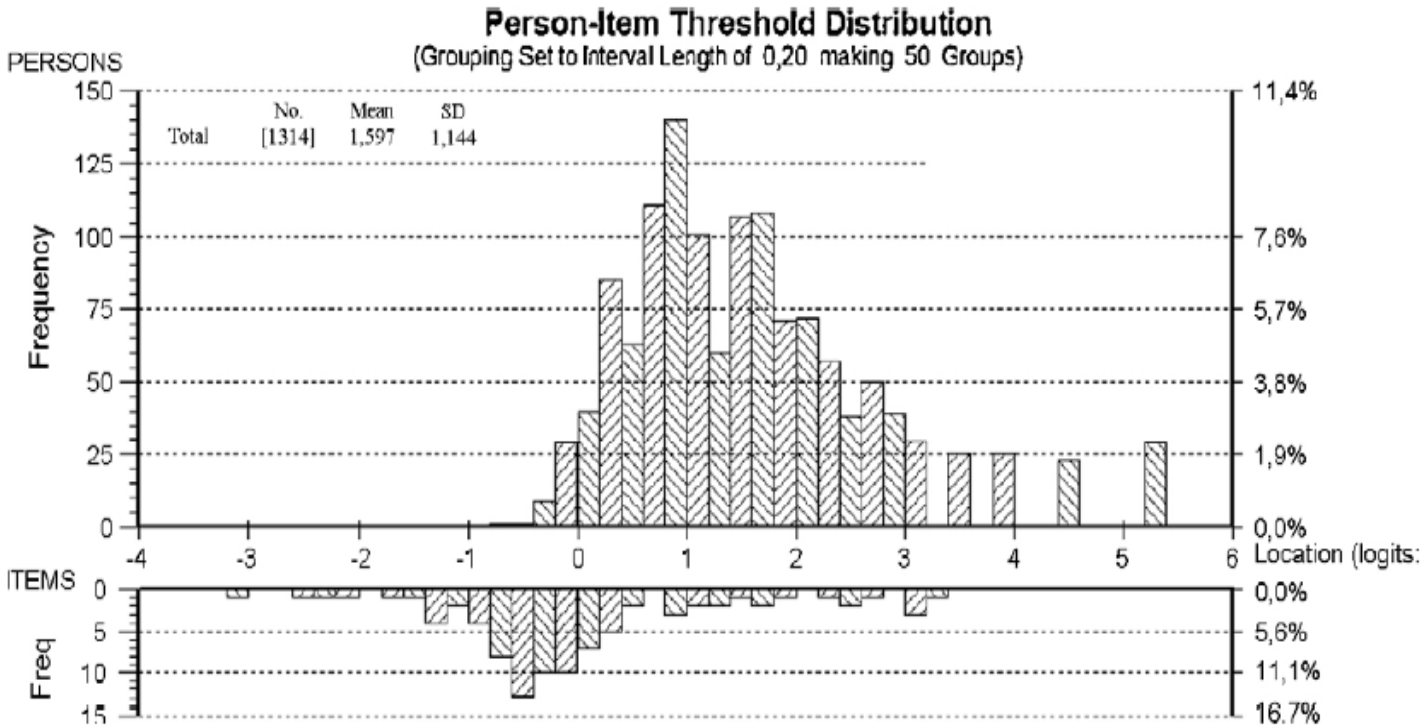
Shifting epistemic practices

Rasch epistemic practices today

- Practitioners using this technique no longer understand the mathematical and statistical intricacies of Rasch analysis.
- Practitioners are now able to deploy Rasch analysis without needing to understand its inner workings.
- There is a new culture of acceptance surrounding Rasch analysis. Researchers working in disparate fields—from clinical medicine to marketing—perform Rasch analysis in the process of scale development.
- Terms such as 'logits', 'thresholds' and 'differential item functioning' are used and understood by a vast array of researchers working in disparate fields.

Shifting epistemic practices

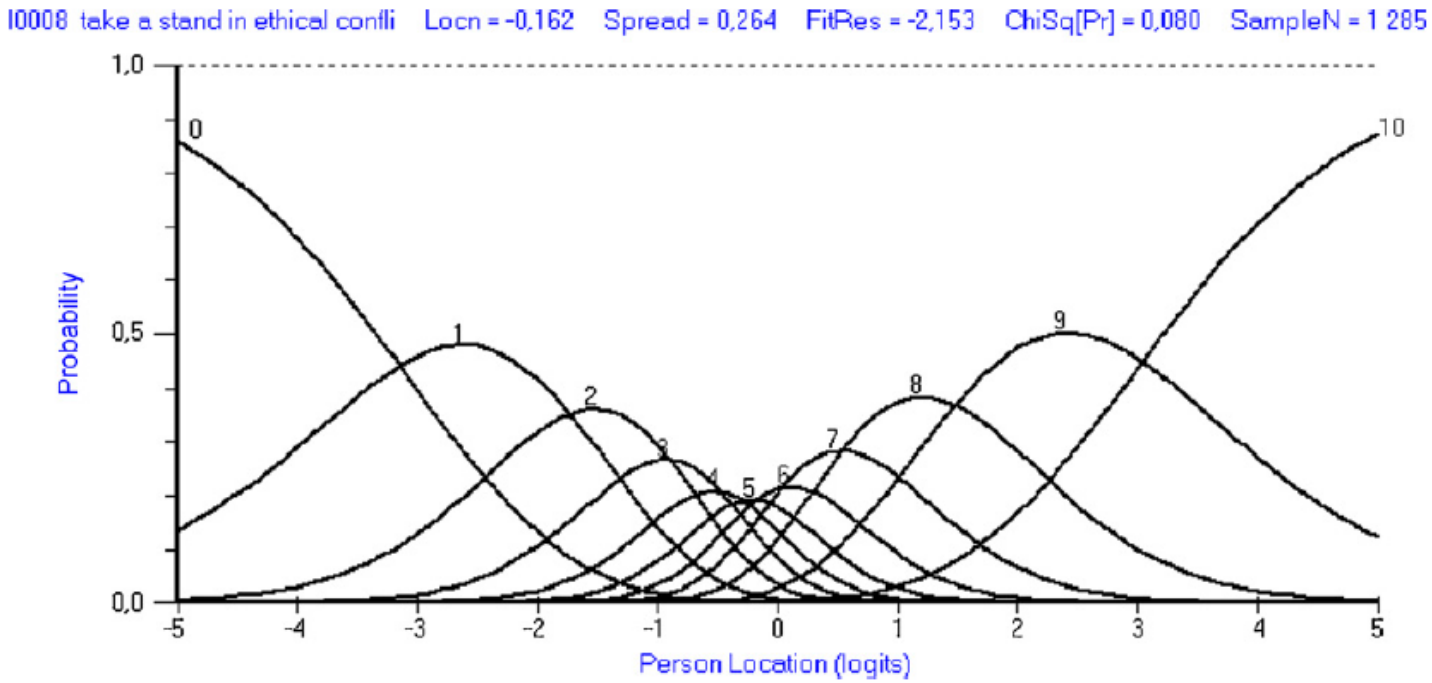
Rasch epistemic practices today



Example of Person-item threshold distribution, by Hagquist, Bruce and Gustavsson.

Shifting epistemic practices

Rasch epistemic practices today



Example of a category probability curve distribution, by Hagquist, Bruce and Gustavsson.

Concluding remarks

- Short courses and training programs on Rasch measurement, and particularly how to use the software packages, are offered frequently around the world [see www.rasch.org].
- Rasch analysis in non-achievement contexts is becoming more and more prevalent as an accepted and fruitful form of epistemic practice.
- Rasch measurement is no longer reserved for the realm of educational testing.
- This analysis makes no claims whatsoever as to the value of Rasch over other techniques. Rather, the focus has been on tracing the shifts in the conceptual pre-conditions, and in epistemic practice, which have influenced its trajectory.

Thank you

Jacob Pearce
Research Fellow
Assessment & Psychometric Research
Jacob.pearce@acer.edu.au