

RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 22 No. 2

Autumn 2008

ISSN 1051-0796

Teaching Students about Rasch Maps

This semester I am teaching a measurement class for graduate students who themselves will teach at K-12 schools. Of course, Bond and Fox is the text-book to help my students learn about measurement so that they may be better informed teachers!

One of the most difficult concepts for my students to master has been how the Wright item-and-person maps are built. They immediately appreciate the way in which the maps can communicate. But they become flustered with the amount of information they have to

synthesize. I had asked my students to analyze the data provided for Chapter 2 of Bond and Fox. That dataset is for an elementary math test (14 children, 12 items), and the shorthand descriptions of the items provided enough information so that my graduate students could understand the items. We discussed the maps in the book, but it was clear to me that many of the students were confused.

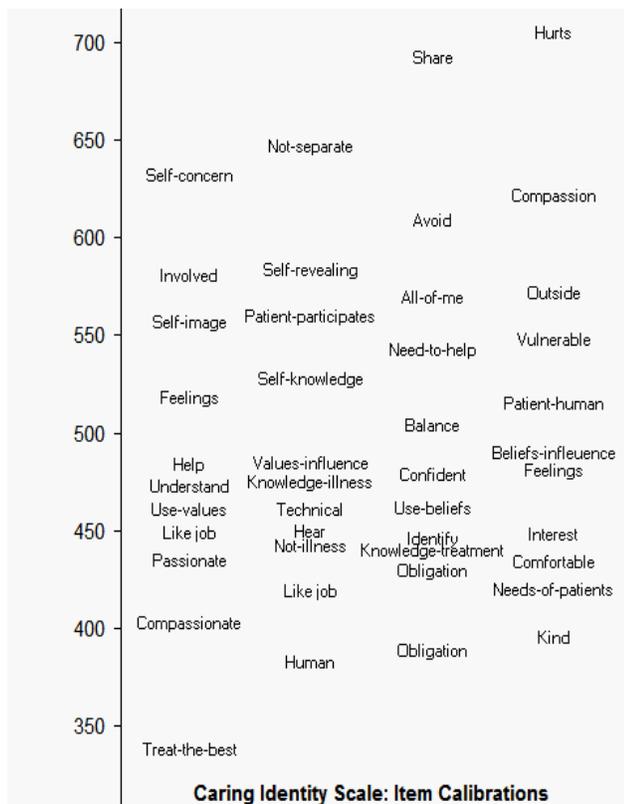
Then I recalled how I learned about Wright maps in stages, so I designed "hands-on" lessons for my students.

First, I provided groups of students with very large sheets of paper (about 1 meter wide and 2 meters long). I explained that we were going to try a step-by-step approach to understanding the Wright maps in Bond and Fox. My students were instructed to draw a vertical line down the left-hand-side of a sheet of paper, and mark out an equal-interval logit scale (4, 3, 2, 1, 0, -1, -2, -3, -4) from top to bottom of the sheet.

They displayed the item-measures, Table 14, in Bond&FoxSteps on their computer screens. Using this Table, they could position each item vertically on the paper, and write down the item's text at that position. Already a map was emerging from the locations of the 12 items.

The next step was to look again at Table 14 and compute the percentage of children who correctly answered the hardest item and the easiest item, writing these percentages next to the relevant item text.

Now we all discussed the ordering and spacing of items. The large sheets were easy for them to read and write upon. The percentages next to the highest and lowest items helped them understand the broad performance of



Wright Item-Map. The item-difficulty hierarchy of the "Caring Identity" scale in Jane F. Sumner & William P. Fisher Jr. (2008) *The Moral Construct of Caring in Nursing as Communicative Action: The Theory and Practice of a Caring Science. Advances in Nursing Science, 31:4, E19-E36*

Table of Contents	
Churchill's paradigm	1165
Cluster-Rasch models	1166
Conceptualizing overfit	1165
Goethal's guideline for item invariance.....	1166
Log-linear Rasch models	1167
Simulating data from marginal scores.	1168
Teaching students about maps	1163

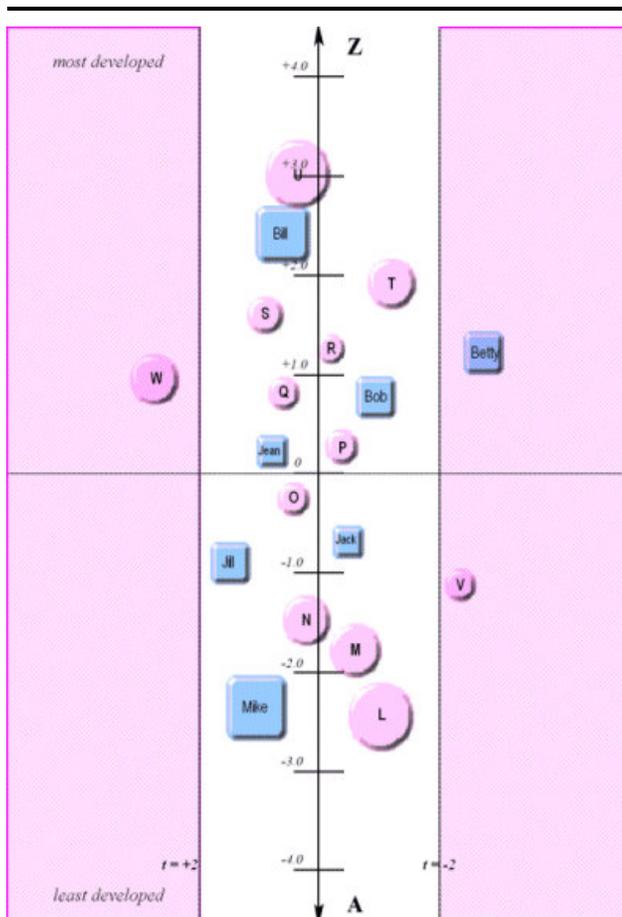
the 14 children.

Following this discussion, another sheet of paper was taped to the left of the first sheet, so that it shared the same vertical logit axis. The students plotted each child's measure using the person-measures from Table 18. I also asked them to compute the success-percentage for the highest and lowest performing children. Just as was done for the easiest and hardest items, the values in percent were noted on the Wright map next to the highest and lowest performing children.

Using this extra-large Wright map, everyone could discuss the strengths and weaknesses of the test. Where were items located? How we could fill in the gaps between the existing items with new items? Did the students perform how we would have expected?

Please spend plenty of time on this activity. As we become familiar with Wright maps we forget what a challenge they are to students new to Rasch. The extra time deepens the students' comprehension of what the measures mean, so speeding up all the learning that follows. No time is lost overall and student involvement increases considerably.

The next phase in the activity involved the use of the



Bond & Fox Pathway Map. Bond T.G. & Fox C.M. (2001) Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Fig. 3.1, p. 22.

large-scale Wright map for the creation of the Bond-and-Fox Pathway map. It is very important to first construct the measure-only Wright map (no horizontal axis), because there is a whole lot more going on measure-and-fit Pathway map! Not only do students need to think about logits, persons, items, and item descriptors, but they must also now think about fit and precision (standard errors).

Each small group of students worked with a sheet of paper about 2 meters wide and 2 meters long. I explained that we would make a Pathway map, using our completed Wright maps as a guide. The students worked out amongst themselves how to use their "old" map to plot the new map. Most students saw that they could use the location of persons and items in the old map to locate persons and items on the new map. And they realized that they now had to add a scale (left to right) on the base of the map to show fit. I helped them think through where they would find the fit data for persons and items (other columns in Tables 14 and 18). I encouraged them to first plot persons and items as dots, ignoring the standard errors of the measures. Physically plotting vertical-scale logit-values for the persons and items, and horizontal-scale fit-values for persons and items was not hard for them to do mechanically. But it was essential to give this task plenty of time, because the students have a lot to think about with the addition of the horizontal fit axis!

Once the students completed the plots, I helped them identify the columns for item error and person error. Then they used those values to draw the precision-circles for the items, and the precision-squares for the persons, centered on the dots they had already plotted. Now they had their own Pathway maps. They understood what they meant, and where they had come from. The maps made good sense. They were no longer mysterious.

Just as Ben Wright would use his yardstick-ruler while he talked, it is helpful to physically create our measuring-instruments in class. This helps Rasch novices better understand and build maps!

William Boone, Miami University (Ohio)
boonewj - at - muohio.edu

ISMTII-2009
The 9th International Symposium on
Measurement Technology and Intelligent Instruments
St. Petersburg, Russia
June 29 - July 2, 2009

Symposium Scope includes:

1. General Problems of Measurement, including:
 - Fundamental Problems
 - Emerging Physical Principles of Measurements
 - Uncertainty, Traceability, Calibration
8. Measurements and Metrology for the Humanitarian Fields
10. Education in Measurement Science

www.tdisie.nsc.ru/ismtii2009

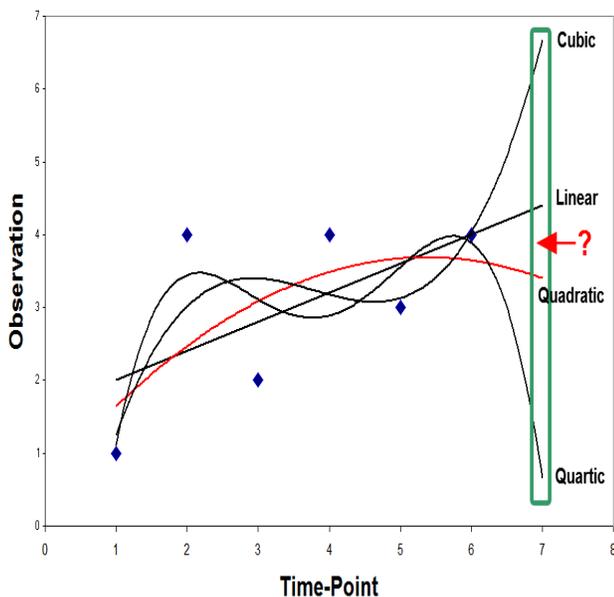
Conceptualizing Overfit

Students stumble over the meaning of “overfit”, indicated in Rasch program output by negative t-statistics and mean-squares less than 1.0 (chi-squares less than their d.f.). This is an aspect of Rasch measurement which sounds strange to analysts familiar with conventional descriptive statistics and model-building.

The nearest thing to overfit in conventional statistical analysis (which fits models to data) is over-parameterization. But statistics courses usually spend little time discussing this problem, so the students imagine that “the better the fit, the better the model!” In building statistical models, they may be instructed to choose between better and worse fitting models, but they are rarely told to reject a model because its fit is “too good”.

A demonstration of over-parameterization is to put some points on an x-y plot, then fit a series of polynomials with more and more terms to those points. It becomes obvious that too many terms actually make prediction of a new point (especially one outside the range of the current points) worse, not better.

Where is the next point?



Here is an example. The observations for 6 time-points are plotted. What is our prediction for the next time-point? The 4 trend-lines are polynomials fitted to the observations. The higher-order is the polynomial, the better is its the fit to the existing points, but, beyond the quadratic, the worse the prediction of the next point.

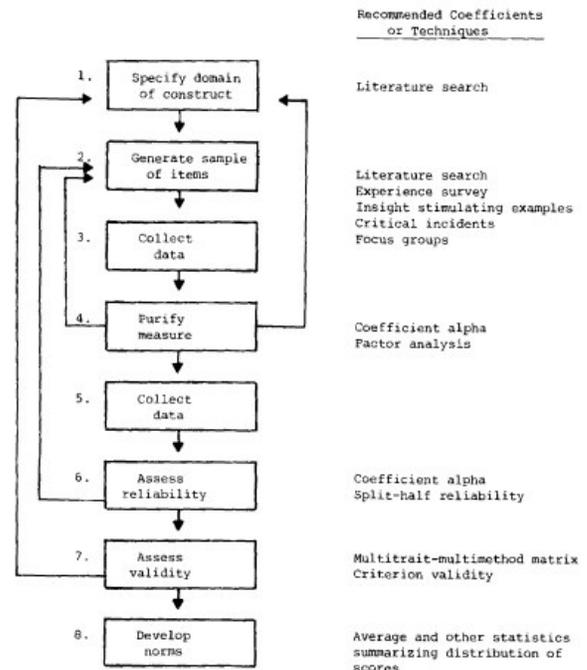
In Rasch, we are fitting data to models, so the Rasch equivalent to overfit due to over-parameterization is overfit due to redundancy (over-predictability) in the data.

Based on a comment by Steve Walter, Graduate Institute of Applied Linguistics.

Churchill's Paradigm for Better Measures

Here is Figure 1 from A Paradigm for Developing Better Measures for Marketing Constructs. Gilbert A. Churchill, Jr., Journal of Marketing Research, 16:1. (Feb., 1979), pp. 64-73).

Figure 1
SUGGESTED PROCEDURE FOR DEVELOPING BETTER MEASURES



A Rasch equivalent to Churchill's raw-score procedure for test construction is explained in detail in Mark Wilson's book "Constructing Measures" (LEA, 2005) www.rasch.org/rmt/rmt183c.htm.

There is also Ben Wright's "The Road to Reason" www.rasch.org/rmt/rmt114o.htm.

An essential difference is that Churchill relies heavily on numerical indicators to improve the measures. Rasch focuses more on the meaning of the measures. Does the hierarchy of the items make sense (construct validity)? Do the measures indicate better performance by persons known to perform better (predictive validity)?

Based on a comment by Juanito Talili.

Research Positions: Australia

The Graduate School of Education at The University of Western Australia has two research positions in Rasch measurement.

www.his.admin.uwa.edu.au/jobvacs/external/academic/ads.htm

The two positions are numbered 2615 and 2616.

Applications close on Feb. 6, 2009.

David Andrich, Ph.D.

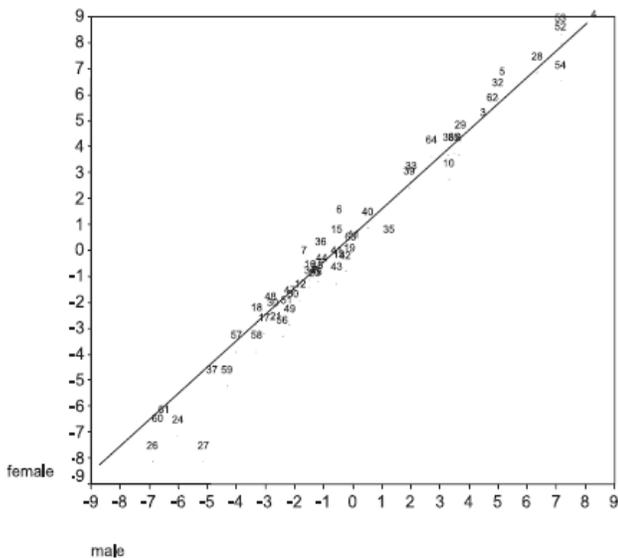
Chapple Chair, Graduate School of Education

The University of Western Australia

Goethals' Guideline for Item Calibration Invariance

The Figure from Kubinger (2005) plots the Rasch item calibrations for 481 males against those for 439 females. Overall, the differences are statistically significant. But Kubinger remarks:

“However, the differences in parameter estimation are of no practical relevance. The differences are too small to impose any systematic handicap on an examinee’s test result because the standard error of estimation of the ability parameter is much greater than errors due to the given item parameter differences. Goethals (1994) provided a rule of thumb: *Any difference of parameter estimations should not exceed a tenth of the range of the parameters.*”



Klaus D. Kubinger (2005) Psychological Test Calibration Using the Rasch Model - Some Critical Suggestions on Traditional Approaches. International Journal of Testing, 5:4, 377 - 394

Goethals, R. (1994). Die praktische Erprobung von Alternativen zur multiple-choice-Vorgabe bei Computertests [Experiences with alternatives for traditional multiple-choice response formats in computerized testing]. Unpublished doctoral dissertation, University of Vienna, Austria.

Rasch Measurement Transactions

P.O. Box 811322, Chicago IL 60681-1322

www.rasch.org/rmt

Editor: John Michael Linacre

Copyright © 2008 Rasch Measurement SIG

Permission to copy is granted.

SIG Chair: Ed Wolfe, Secretary: Timothy Muckle

Program Chairs: Dimiter Dimitrov & Diana Bernbaum

SIG website: www.raschsig.org/

Cluster-Rasch Models

Li and Hong (L&H, 2001) propose a multi-stage model for relating gene expression profiles to phenotypes. It estimates several Rasch measures associated with each tissue-sample (“person”) based on its gene (“item”) expression (“response”) profile. These measures are used to predict the phenotypes (physical characteristics).

Stage 1: Each data-point is the frequency of a gene-indicator in a tissue sample. The frequencies are stratified into ordinal responses. L&H choose 4 strata, so that the ordinal responses become 0,1,2,3. In their examples, there are about 100 genes and 72 tissue-samples.

Stage 2: The genes are segmented into clusters. L&H use the original data-points, but better would be the residuals from a Rasch analysis of the entire dataset.

Stage 3: A Rasch Partial-Credit analysis is performed on each gene-cluster. This constructs a latent unidimensional variable on which each gene and each tissue-sample is located. The measure for each tissue-sample for each gene cluster is estimated. L&H model each gene to have its own rating-scale structure, but they note that their data are somewhat thin for this type of analysis. A Rasch-Andrich (or gene-grouped) model would be more robust.

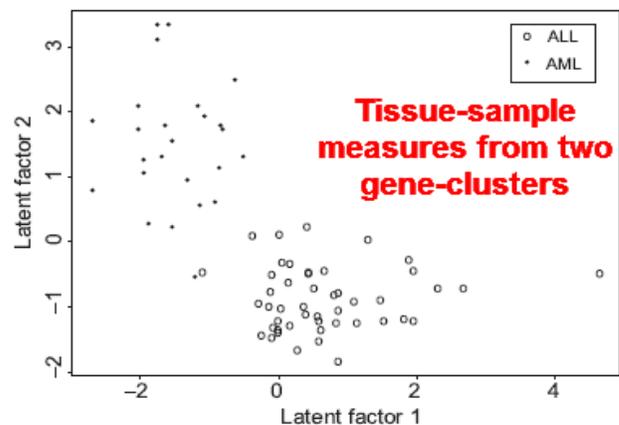
Stage 4: The cluster-measures for each tissue-sample are regressed on its observed phenotype, producing a set of regression parameters.

Stage 5: The phenotype of a new tissue-sample is predicted based on its gene expression, the clustering of the genes, and the regression parameters. The predicted phenotype is compared with its observed phenotype to evaluate the success of the cluster-Rasch model.

The Figure cross-plots the measures (latent factors) for two groups of tissue-sample and two clusters of genes.

Li H. & Hong F. (2001). Cluster-Rasch models for microarray gene expression data. Genome Biology, 2(8): research0031.1–0031.13

<http://genomebiology.com/2001/2/8/research/0031.1>



Advances in Log-Linear Rasch Models

The ever-increasing complexity of Rasch datasets, combined with a desire to use standard statistical software for estimation has motivated advances in log-linear Rasch models. The relationship between log-linear Rasch models and logit-linear (or exponential) Rasch models is shown at www.rasch.org/rmt/rmt113r.htm

Log-linear models simplify estimation by eliminating nuisance parameters (usually those of the subjects), but often add complexity through the need for design matrices. They are also awkward to implement when data-points are missing.

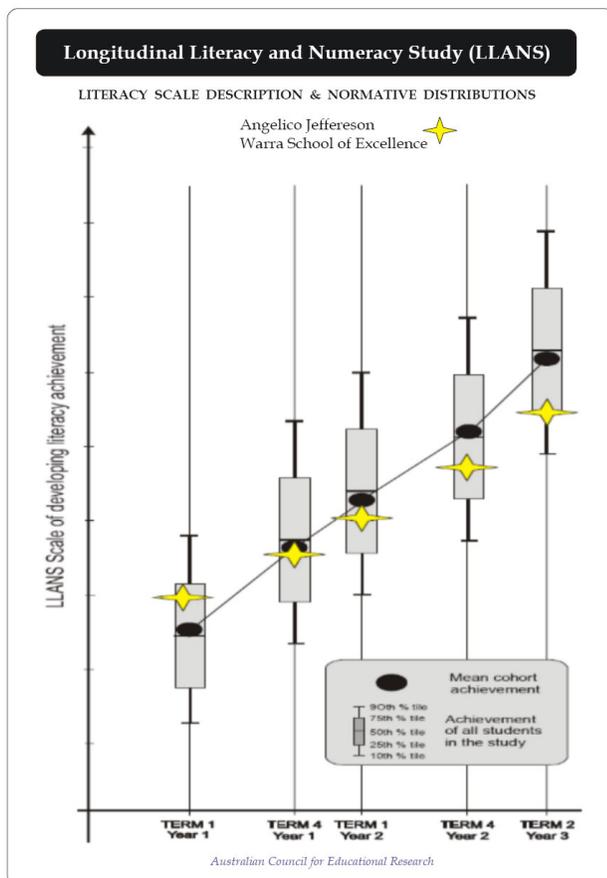
Hatzinger & Katzenbeisser (2008) derive dichotomous and partial-credit log-linear Rasch models incorporating multiple time-points. This can be estimated using conditional maximum-likelihood estimation (CMLE) with standard statistical software, such as R. It models dependency across time-points, and also allows different subjects to be observed at different numbers of time-points. The example dataset has dichotomous data, 3 items, 45 subjects observed at up to 11 time-points.

The estimation of the parameters of multidimensional polytomous Rasch models presents an even greater technical challenge. Anderson et al. (2007) achieve it by formulating the Rasch model as a log-linear-by-association (LLA) model. The multidimensional structure renders conventional CMLE impossible in general, so a pseudo-likelihood technique is employed. The overall likelihood of the data is decomposed into a set of parallel regression models which are maximized simultaneously. This is implemented in the plRasch package for R, and the SAS plgRasch macro. Example datasets have up to 30 items, 1000 subjects, 3 response categories and 2 dimensions.

Anderson, C.J., Li, Z., & Vermunt, J.K. (2007). Estimation of models in the Rasch family for polytomous items and multiple latent variables. *Journal of Statistical Software*, 20:4. www.jstatsoft.org/v20/i06

Hatzinger R. & Katzenbeisser W. (2008). Log-linear Rasch-type Models for Repeated Categorical Data with a Psychobiological Application. Department of Statistics and Mathematics, Wirtschaftsuniversitaet Wien. Research Report 69.

http://epub.wu-wien.ac.at/dyn/virlib/wp/eng/showentry?ID=epub-wu-01_dab



A growth map of achievement progress in literacy showing individual and norm-referenced growth against descriptions of domain-referenced criteria. *K.J. Rowe (2006) Assessment during the early and middle years: Getting the basics right. ACER.* http://www.acer.edu.au/documents/LP_Rowe-AssessmentintheearlyandmiddleyearsJuly2006.pdf

“Improving Efficiency in Health Outcome Measurement.” Rehabilitation Institute of Chicago Saturday, March 28, 2009.

The goal of the conference is to provide health and rehabilitation administrators, clinicians and researchers with state-of-the-art advances in objective measurement as it relates to health care outcomes.

9:00 AM – 4:30 PM at the Rehabilitation Institute of Chicago, 345 E. Superior St. , Chicago IL 60611 in the Magnuson Auditorium.

www.ric.org/research/centers/cror/projects/rrtc/data/T3.aspx

Topics to be presented include:

The role of theory in generating items and instruments for health and rehabilitation outcomes research.

Techniques for improving instruments.

Options for data collection and their application in health care settings.

The use of IRT/Rasch and computer adaptive testing (CAT) to streamline collection of patient-reported outcomes.

Confirmed speakers include: **Benjamin D. Wright**, Ed Bouchard, David Cella, William Fisher, Jr., Richard Gershon, Carl Granger, Dennis Hart, Jin-Shei Lai , Robert W. Massof, A. Jackson Stenner, David Tulskey.

Simulating Data from Marginal Scores

Simulating Rasch datasets from known item difficulties and person abilities is straightforward www.rasch.org/rmt/rmt213a.htm

More challenging is simulating Rasch datasets with known marginal scores. Verhelst et al. (2007) achieve this for complete rectangular dichotomous datasets using a Markov-Chain Monte-Carlo (MCMC) algorithm. Its operates on dichotomous data by shuffling observations between rows and columns. The program is implemented as an R package, RaschSampler, which can simulate matrices of up to 1024 rows and 64 columns.

Another approach is to simulate the data using Rasch measures. The dataset can be dichotomous or polytomous, complete or incomplete, and of any size.

1. Compute the marginal scores (persons, items) and counts (polytomous categories) in the generating dataset.

2. In the simulated dataset, impute the extreme observations corresponding to all extreme marginal scores in the generating dataset. Flag those observations as missing in the generating dataset. If there were extreme marginal scores, repeat from 1.

3. Estimate the measures (persons, items, polytomies) for the generating dataset. Rough estimates are good enough. Polytomies: allow for intermediate categories with sampling zeroes.

4. Select a non-missing observation at random in the generating dataset. Simulate its value based on the estimated measures and place it in the matching location in the simulated dataset. Flag that data-point as missing in the generating dataset.

5. Repeat the procedure from 1 until there are no active observations in the original dataset. For dichotomous data, only the measures for one person and one item will need to be re-estimated for each simulated observation. For polytomous data all measures may need to be re-estimated.

When this procedure completes, the simulated dataset will have the same marginal scores and category counts as the original generating dataset.

John Michael Linacre

Verhelst N., Hatzinger R., Mair R. (2007) The Rasch Sampler. Journal of Statistical Software, 20:6. <http://www.jstatsoft.org/v20/i04>

Rasch-related Coming Events

- Jan. 2-30, 2009, Fri.-Fri. Practical Rasch Measurement - Core Topics online course (M. Linacre, Winsteps), www.statistics.com
- Feb. 3, 2009, Tues. 4th Meeting of the UK Rasch Users Group, Cambridge, UK www.rasch.org.uk/
- Feb. 27 - March 27, 2009, Fri.-Fri. Rasch - Further Topics online course (M. Linacre, Winsteps), www.statistics.com
- March 11-13, 2009, Wed.-Fri. Introduction to Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric
- March 15, 2009, Sun. Online Programs in Research Methodology (Fall 2009, registration deadline), www.rasch.org/onlineuic.htm
- March 28, 2009, Sat. Symposium: Improving Efficiency in Health Outcome Measurement, Chicago www.ric.org
- April 13-17, 2009, Mon.-Fri. AERA Annual Meeting, San Diego, CA, USA, www.aera.net
- May 1-29, 2009, Fri.-Fri. Many-Facet Rasch Measurement online course (M. Linacre, Facets), www.statistics.com
- May 13-15, 2009, Wed.-Fri. Introduction to Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric
- May 18-20, 2009, Mon.-Wed. Intermediate Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric
- June 2-3, 2009, Tues.-Wed. 2009 GMAC Invitational Conference on Computerized Adaptive Testing, Minneapolis, www.gmac.com/CATConference
- June 26 - July 24, 2009, Fri.-Fri. Rasch - Core Topics online course (M. Linacre, Winsteps), www.statistics.com
- June 29 - July 2, 2009, Mon.-Thur. 9th International Symposium on Measurement Technology and Intelligent Instruments, Russia, www.tdisie.nsc.ru/ismtii2009
- July 28-30, 2009, Tues.-Thur. PROMS HK 2009 Pacific Rim Objective Measurement Symposium, Hong Kong
- Aug. 21 - Sept. 18, 2009, Fri.-Fri. Rasch - Further Topics online course (M. Linacre, Winsteps), www.statistics.com
- Sept. 6-11, 2009, Sun.-Fri. IMEKO XIX World Congress: Fundamental and Applied Metrology, Portugal www.imeko.org
- August, 2010 Probabilistic models for Measurement - 50 years, Conference, Copenhagen, Denmark