RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 28 No. 1 | Summer 2014 | ISSN 1051-0796

Reflections of IOMW 2014

Dense, imaginative, wide-ranging, passionate, international -- IOMW 2014, held last April in Philadelphia and featuring over 60 presentations, was an example of the Rasch measurement community in top form.

After a day of workshops on RUMM, Winsteps, ConQuest, Facets, and Damon on Python, Jack Stenner, CEO of MetaMetrics, set the theme with his critique of the Cronbach, Messick, Kane perspective on validity and the profusion of "validities" afflicting the psychometric world. Michael Kane, the Messick Chair at ETS, was on hand to provide an elegantly crafted counter-perspective. Andrew Maul, David Torres Irribarra, and Joshua McGrane explored the philosophical foundations and epistemology of measurement, followed by practical examples in writing by Nadia Behizadeh and George Engelhard and a lovely encapsulation by Brent Duckor, the architect of the "Foundations" session.

That was the warm-up.

The session on New Approaches and Disciplines featured a thought-provoking section on metrology, medical outcomes, and social variables organized by William Fisher who brought in bridge researchers like Leslie Pendrill, Stefan Cano, Jeremy Hobart, and Robert Massof. The idea of merging Rasch models with the science of metrology -- establishing measurement standards across fields -- is particularly intriguing and amply demonstrates the need for Rasch-like objectivity requirements in the health and social sciences. There were presentations on the linear logistic test model (LLTM) and Q-Matrices led by Lin Ma and Kathy Green. George Engelhard and Jue Wang presented Erma, an R package for Rasch paired comparison models. Stefanie Wind explored the use of distractor information. Robert Massof offered an ingenious new way to model guessing.

The second day morning session on Applications and Modeling led with David Andrich's spirited defense of the partial credit model as a Rasch model, followed by presentations on threshold ordering and positive and negative affect. Another series of papers organized by William Fisher and the "Berkeley school" explored how to use constructs to create "trading zones" for the translation of information across communities using the BEAR Assessment System (BAS). Concurrent sessions explored factors that cause artifactual DIF (Curt Hagquist), models for evaluating rater accuracy (Ed Wolfe), and evaluation of equal-interval properties on vertical scales (Derek Briggs). Massof showed how the Medicare system might benefit from Rasch modeling to set the "G-codes" used to categorize patient functional status.

Another series of papers shared how the Rasch model is being used in Saskatchewan to measure oral language development through story-telling (Patrick Charles, Michelle Belisle), in British Columbia to measure psychiatric recovery and analytic rumination (Skye Barbic), in Singapore to measure critical thinking skills (Raymond Fong), in Korea to scale the National English Ability Test (Seul Ki Koo, Yongsang Lee), and in Quebec to measure the efficacy of peer assessment (Christophe Chénier, Nadine Talbot). Maria Veronica Santelices from Chile discussed her ground-breaking work on relating teacher quality to teacher and school characteristics.

Table of Contents

The second day afternoon session was dedicated to multidimensionality. One series of papers explored questions related to ConQuest's Multidimensional Random Coefficient Multinomial Logit model. Hanke Korpershoek explored the multidimensionality of school motivation in the Netherlands. Nathaniel Brown examined heteroscedastic interactions between constructs. Steffen Brandt offered remarkable insights on the interpretation of unidimensional measures derived from multidimensional data.

The second series of multidimensionality papers responded to the "2014 IOMW Subscale Measurement Challenge". A bottle of cold Danish Carlsberg beer was awarded to the contestant who could most accurately predict subscale measures derived from a three-construct language test administered in Iran. Mark Moulton and Gregory Stone presented the NOUS model for highly multidimensional data and showed how it could be used to trade information between subspaces. Massof offered an insightful analysis based on factor analysis. Michael Linacre demonstrated a clever use of regression and the many-facets model for computing subscale measures, but alas, fell foul of the contest rules by using information from the validation set. It was Edward Wolfe who won the competition by two percentage points using a two-dimensional ConQuest-based model, whipped out in an afternoon. The subscale challenge proved to be highly interesting and gets at under-appreciated psychometric issues arising from the development of new item types for Common Core.

Courtney Tobiassen at University of Denver was awarded the Graduate Student Best Paper award for her paper, "Differential Item Functioning on a Measure of Perceptions of Preparation for Teachers, Teacher Candidates, and Program Personnel." Papers and abstracts for IOMW 2014 can be obtained through the IOMW website: www.iomw.org.

The next IOMW will be held April 2016 in Washington, DC. We expect it to address, among other things, the problem of "big data" and how to incorporate the latest developments in data science.

Appreciation is extended to the IOMW Conference Committee for putting on a splendid event: Brent Duckor, William Fisher, Gregory Stone, Jade Caines, Stefanie Sebok, and Mark Moulton. Richard Smith was deeply involved in the early stages of organization. Anyone interested in participating in the work of IOMW should contact Mark Moulton at iomwconfcomm@gmail.com.

Mark Moulton

Pictured here: Michael Kane

Pictured here: David Andrich

Pictured here: Mark Wilson

Rasch Measurement Transactions
www.rasch.org/rmt
Editor: Kenneth Royal
Email Submissions to: Editor \at/ Rasch.org

RMT Editor Emeritus: John M. Linacre
Rasch SIG Chair: Vacant
Secretary: Mikaela Raddatz
Treasurer: Leigh Williams
Program Chairs: Jessica Cunningham& Sara Hennings
Rasch SIG website: www.raschsig.org

Pictured here: Robert Massof

Pictured here: William Fisher and Leslie Pendrill

Rasch SIG Business Meeting

David Andrich provided the annual SIG Business Meeting presentation. His talk was entitled "On Rasch Measurement Theory". Pictured below is Tim O'Neil, Rasch SIG Chair, and David Andrich

Rasch Estimates for Standard Datasets

Let's build a library of standard datasets and their Rasch estimates. These can be used to confirm that Rasch software is functioning correctly and also for teaching about Rasch estimation.

Estimation method:
AMLE = Anchored Maximum Likelihood Estimation (MLE for estimating person abilities with known item difficulties)
CMLE = Conditional Maximum Likelihood Estimation (R- eRm, WINMIRA)
JMLE = Joint Maximum Likelihood Estimation (R-mixRasch, Winsteps)
MMLE = Marginal Maximum Likelihood Estimation (R-ltm, ConQuest)
PMLE = Pairwise Maximum Likelihood Estimation (R-pairwise, RUMM2030)
WMLE = Warm's Mean Likelihood Estimation (applied to MLE estimates)

All estimates are in logits. The estimate for column 1 (item 1) is set to 0.0 logits.

Standard dataset 1:
Complete dichotomous dataset of 2 columns (items) and 2 rows (persons):
0,1
1,0
All Rasch estimation methods: column estimates: 0.0, 0.0 ; row estimates: 0.0, 0.0.

Standard dataset 2:
Complete dichotomous dataset of 2 columns (items) and 3 rows (persons):
0,1
1,0
0,1
CMLE column estimates: 0.00000, -0.69315
AMLE row estimates: -0.34658, -0.34658, -0.34658

JMLE column estimates: 0.00000, -1.38629
JMLE row estimates: -0.69315, -0.69315, -0.69315

MMLE column estimates: 0.00000, -1.38629
AMLE row estimates: -0.69315, -0.69315, -0.69315

PMLE column estimates: 0.00000, -0.69315
AMLE row estimates: -0.34658, -0.34658, -0.34658

Standard dataset 3:
Complete dichotomous dataset of 3 columns (items) and 3 rows (persons):
1,0,0
0,1,1
0,1,1

CMLE column estimates: 0.00000, -1.00505, -1.00505
AMLE row estimates: -1.40449, 0.05635, 0.05635
WMLE row estimates: -1.17098, -0.18498, -0.18498

JMLE column estimates: 0.00000, -1.56593, -1.56593
JMLE row estimates: -1.84142, -0.27549, -0.27549
WMLE row estimates: -1.63506, -0.50850, -0.50850

MMLE column estimates: 0.00000, -1.38629, -1.38629
AMLE row estimates: -1.69820, -0.17070, -0.17070
WMLE row estimates: -1.48169, -0.40644, -0.40644

PMLE column estimates: 0.00000, -0.69315, -0.69315
AMLE row estimates: -1.17436, 0.24746, 0.24746
WMLE row estimates: -0.93166, 0.00210, 0.00210

Comments, corrections and suggestions for more standard datasets are welcome.

John Michael Linacre
mike@winsteps.com

Causing and Being Caused: Items in a Questionnaire May Play a Different Role, Depending on the Complexity of the Variable

I meditated on two nice, enlightening articles that appeared in RMT, 22:1 and 22:4 (2008 and 2009, respectively). The articles, written by Stenner, Stone and Burdick concerned the causal vs. correlational relationship between indicators and variables. Basically, there are variables ("index" variables) that *are caused* by their indicators (indexes) and variables (reflective) that *cause* their observable indicators (measurement indicators). In a questionnaire, "formative" items do not generate a truly "latent" variable: they are the only game in town and their result is entirely observable. They constitute a checklist more than a true measure. "Reflective" items, by contrast, do shed some light on the latent variable, and provide (an estimate of) a true measure. Corollaries of this epistemic approach are that a) formative variables are artifactual (hence, dangerous) constructs, whereas true "latent" variables do exist whichever their indicators, and b) an ideal item-response scale should be formed by "reflective" indicators as opposed to "formative" indicators.

I entirely agree with the example of the variable "socioeconomic status" given in the former article, where "education", "income", etc., really are causes of SES rather than a selection of "reflective" items. Elsewhere, I highlighted the risks of using such artifactual constructs (specifically, Quality of Life) for concealing political decisions on the rationing of healthcare resources under the guise of "objective measurement" (Tesio, 2009). By contrast, I found debatable an example given in the second Stenner article, suggesting a wrong interpretation of the FIM as a measure, whereas it should actually be considered an "index". Let's cite the text:

"The Rasch model has been shown to fit FIM data reasonably well, which indicates that the scale locations describe adequately the relative order in which these functions are lost in the aging population. The items on the top describe difficult activities, such as climbing stairs, whereas items on the bottom describe easier activities that are maintained relatively well. (Embretson, 2006, p. 52).

Contrary to a latent variable interpretation, the FIM (Functional Independence Measure) appears to be an index of motor functioning with the causal action moving from indicators to index. If the desired medical outcome is "more functional independence," then rehabilitating bladder control, walking, bathing, and so on should promote the intended outcome rather than the other way around. Alternatively, we could teach the patient to drive a motorized wheelchair but to include this as an indicator would alter the definition of "functional independence".

I think that the FIM provides evidence of the fact that being an "index" rather than a "measure" is not necessarily an all-or-nothing concept (do such phenomena really exist on this planet?). It is true that doing effective rehabilitation exercise focused on a given item (e.g. walking) does not coax the other items towards similar improvement. In a paper of mine aimed at developing a scale of balance in multiple sclerosis patients (Tesio et al., 1997) I evidenced a trouble in the final instrument, namely in the Rasch glossary, a Differential Item Functioning (DIF) between pre- and post-treatment item calibrations. Perhaps this was not a trouble in the scale, but in the treatment! My interpretation was that traditional balance training is too focused on "resistance to external perturbations", while "resistance to self-perturbations" is relatively overlooked: hence the differential changes in relative item difficulty. Going back to the walking example, "rehabilitation of walking" is an assortment of behavioral interventions entailing stimulation of balance, force, attention, motivation, communication etc., given that it is a teaching, highly relational activity. It is reasonable to conceive that any improvement in "walking" is indeed associated with some improvement (only some, of course) in the rest of the indicators.

The articles by Stenner and colleagues made me reflect on the fact that we cannot treat "independence in daily life" *per se*, although it is the goal of our work: we can only treat "indicators", such as continence, speech, balance, etc. My feeling is that all these can be envisaged as lying on different locations along a continuum spanning between the extreme roles of "formative" vs. "reflective" variables. For instance, in the FIM scale "bladder continence" can be altered irrespective of many other cognitive and behavioral attributes (imagine a young cognitively intact paraplegic): and in fact it is prone to relevant DIF across diagnostic classes. By contrast, "lower body dressing" implies motor and sensory skills, cognition, motivation, social relationships (why dressing the lower body if not for out-of-bed mobility and social interaction?), and it is much less prone to DIF across diagnostic classes. Of course, the more we manipulate (e.g. by treatment) a "reflective" indicator, the more we can assume we are manipulating all of the other indicators, and thus we can hope that change in the target indicator will "reflect" a change of the whole variable (and will be correlated with changes in the whole item set). I suspect that the more an indicator can be assumed to belong to the person as a whole (let me call it a *high-order behavioral indicator*, see below), as opposed to body parts or focal functions, the more it is *reflective* of the latent person's variable. Thus, in principle, interventions on reflective indicators are preferable. However, at least in physical and rehabilitation medicine, this raises the risk of aiming at purely "adaptive" outcomes: if the goal is "independence in daily life" after stroke, an awkward spastic gait may appear to be an outcome equivalent to a more physiologic gait, so why bother with more fine-tuned training? The latter might require work to be focused upon highly local phenomena (such as, say, passive mobility of the ankle, knee joint kinematics etc.) which would appear as roughly "formative" once added as items to the FIM. My objections are:

1. A person cannot be described by just one variable (e.g., what about "satisfaction with the outcome"? And what about "risk for fall"?). People, not statistics, must decide what variables represent the goals of treatment.
2. Latent "persons" variables are not only multiple (potentially infinite?) but can also be thought of as located along a gradient spanning from less-to-more complexity of behaviors and perceptions (Tesio, 2003). By complexity (literally, from the Latin, *cum-plexus*, "interwoven") I mean here the number and the order of interactions across "simpler" person's traits, allowing for the trait of interest. For instance, "balance" can be thought of as of lower "order", compared to "independence in daily life": the latter implies the former, not the reverse).
3. There is a complex non-linear liaison between biological (referred to "parts" of the body) and behavioral variables (referred to a unitary "person") (Granger & Linn, 2000; Tesio, 2004). As the ancient Romans said, one should distinguish between risks "quoad vitam" (threats to life) and those "quoad valetudinem" (threats to "ability"). In fact, all living beings adapt to biological troubles in order to restore behavioral competence. People, however, are unique in that they can also treat biological problems, thus aiming at "intrinsic", rather than only "adaptive" recovery.

If my objections hold, an indicator that appears to be "formative" with respect to a high-order variable, can be "reflective" with respect to a lower-order one, closer to the biological extreme. Joint pain may be "formative" (hence, a poor item) with respect to "independence in daily life", but "reflective" with respect to "perceived effectiveness of an anti-inflammatory drug".

Luigi Tesio

References

Granger C.V., & Linn R.T. (2000). Biologic patterns of disability. *Journal of Outcome Measurement, 4*(2):595-615.

Stenner J., Burdick D.S., & Stone M.H. (2008). Formative and reflective models: can a Rasch Analysis tell the difference? *Rasch Measurement Transactions, 22*(1), 1152-53

Stenner J., Stone M.H., & Burdick D.S. (2009). Indexing vs. measuring. *Rasch Measurement Transactions*, *22*(4),1176-77.

Tesio L. (2009). Quality of life measurement: one size fits all. Rehabilitation makes no exception. *Journal of Medicine and the Person*,*7*:5-9.

Tesio L., Franchignoni F.P., Battaglia M.A., & Perucca L.(1997). A short measure of balance in Multiple sclerosis: validation through Rasch analysis. *Functional Neurology*, *12*(5):255-65.

Tesio L. (2003). Measuring person's behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine*,*35*:1-11.

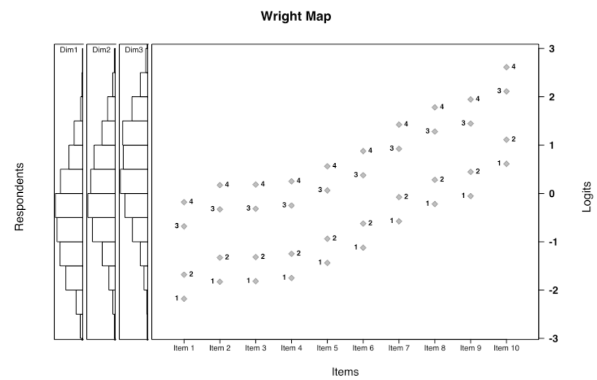
Tesio L. (2004). Bridging the gap between Biology and Clinical Medicine. Some help from Rasch measurement theory. *Journal of Applied Measurement*, *5*(4):362-366.

Journal of Applied Measurement
Vol. 15, No. 2, 2014

Richard M. Smith, Editor, www.jampress.org

Wright Map Package for R

A powerful yet simple graphical tool available in the field of psychometrics is the Wright Map (named after Ben Wright), which presents the location of both respondents and items on the same scale.



The WrightMap package provides functions to easily create these beautiful Wright Maps from item parameters and person estimates stored as R objects. The plots can represent polytomous and multidimensional models, are highly customizable, and, as any other R plot, can be exported into multiple image formats.

Although the package can be used in conjunction with any software used to estimate the IRT model (e.g. eRm or IRToys in R, or Stata, Mplus, etc.), WrightMap features special integration with ConQuest to facilitate reading and plotting of its output directly.

You can find tutorials and information about the package on the official website wrightmap.org or download it from the WrightMap page on CRAN. If you can any questions, you can reach us by going to the wrightmap.org/ask page.

David Torres Irribarra & Rebecca Freund
University of California - Berkeley

Notable Quote

A certain amount of opposition is a great help to a man.
Kites rise against, not with the wind.
-John Neal, 1793-1876

A Pairwise Algorithm in R for Rater-Mediated Assessments

Rasch measurement theory is fundamentally based on the concept of comparisons. Rasch (1977) argued that all scientific statements "deal with comparisons, and the comparisons should be objective" (p. 68). There are a variety of algorithms for estimating the parameters of Rasch measurement models, but none has the didactic value and simple elegance of the pairwise algorithm. Rasch (1960/1980) attributed an early pairwise algorithm to Leunbach (pp. 171-172). Operational versions of several pairwise algorithms were developed by Choppin (1968, 1985, 1987), and extended by Garner and Engelhard (2009). We are currently working on an R program (Wang & Engelhard, 2014) that can be used to estimate the parameters of the Partial Credit (Masters, 1982), Rating Scale (Andrich, 1978) and Many Facet (Linacre, 1989) Rasch models based on the pairwise algorithm.

One of the advantages of pairwise approach is that a simple counting algorithm can be used to illustrate the idea of invariant comparisons and objectivity that is fundamental to Rasch measurement theory (Rasch, 1977; Engelhard, 2013). The use of R software (R Core Team, 2014) gives everyone a chance to use Rasch measurement models with free and open source code. The basic counting algorithm underlying the pairwise algorithm, as well as least square estimates of generalized item locations, can be written simply in R code:

```
x <- G
Nperson <- dim(x)[1]
Nitem <- dim(x)[2]
B <- matrix(0, Nitem, Nitem)
for(k in 1:Nperson){
for(i in 1:Nitem){
for(j in 1:Nitem){
if(is.na(x[k,i])==FALSE &
is.na(x[k,j])==FALSE){
if(x[k,i]>x[k,j]){B[i,j] <- B[i,j]+1} } } } }
B2 <- B %*% B
for(i in 1: dim(B)[1]-1){
if (min(B2, na.rm=TRUE)==0){
B2 <- B2 %*% B } }
D <- t(B2)/B2
Logit <- log(D)
```

In order to illustrate the algorithm, Table 1 reports the ratings from a single rater taken from Garner and Engelhard (2009), and Table 2 shows the Guttman recoded generalized items (G matrix) based on the ordered ratings. Essentially, each ordered rating is conceptualized as a generalized item scored as separate dichotomous items. For example, a rating of 0 is recoded as [0 0], a rating of 1 is recoded as [1 0], and a rating of 2 as [1 1]. We use this Guttman recoded summary of polytomous ratings in our pairwise comparison algorithm.

Table 1
One rater rating 10 persons on three items (Garner & Engelhard, 2009)

Rater:	**1**	**1**	**1**
Item:	**1**	**2**	**3**
Person			
1	2	2	1
2	2	1	2
3	1	0	0
4	0	1	1
5	1	2	2
6	2	2	0
7	2	2	1
8	1	0	1
9	2	2	1
10	2	1	0

Table 2
Recoded Guttman (G) matrix with generalized items

G-Item:	**1**	**2**	**3**	**4**	**5**	**6**
Rater:	**1**	**1**	**1**	**1**	**1**	**1**
Item:	**1**	**1**	**1**	**1**	**1**	**1**
Threshold:	**1**	**2**	**1**	**2**	**1**	**2**
Person						
1	1	1	1	1	1	0
2	1	1	1	0	1	1
3	1	0	0	0	0	0
4	0	0	1	0	1	0
5	1	0	1	1	1	1
6	1	1	1	1	0	0
7	1	1	1	1	1	0
8	1	0	0	0	1	0
9	1	1	1	1	1	0
10	1	1	1	0	0	0

According to Choppin (1968, 1985), if the total score of individuals is one on item i and j, b_{ij} of them get item i correct and item j wrong, and b_{ji} of them get item j correct and item i wrong, then a paired comparison matrix B can be created. The B matrix is an adjacency matrix that can be constructed with entries b_{ij}, and the transpose of matrix B is composed of entries b_{ji}. For example, there are three people who got generalized item 1 correct but generalized item 2 wrong based on the Guttman recoded data; then the cell of the first row and the second column, which defines b_{12}, is 3. Garner and Engelhard (2002) recognized that powers of the adjacency matrix (B) can be used to increase connectivity through indirect comparisons. For the example data in this study, the B matrix was raised to the third power (B^3) in order to eliminate zeroes in the adjacency matrix, which means that all of the generalized items have direct or indirect connections with each other.

Table 3
Adjacency (B) matrix for six generalized items

G-Items	1	2	3	4	5	6
1	0	3	2	4	3	7
2	0	0	0	2	2	5
3	1	2	0	3	2	6
4	0	1	0	0	1	4
5	1	3	1	3	0	5
6	0	1	0	1	0	0

Table 4
Adjacency (B^3) matrix for six generalized items

G-Items	1	2	3	4	5	6
1	17	149	24	177	97	355
2	4	45	6	64	47	145
3	14	103	14	125	70	258
4	3	30	4	37	28	94
5	16	107	16	118	51	226
6	3	20	3	20	4	28

The D matrix (Table 5) is obtained by dividing the B^3 matrix by the transpose of the B^3 matrix, which is the ratio of b_{ij}/ b_{ji}. By taking the logarithm of these ratios, a Logit matrix (Table 6) is created with all the relative difficulties between each pair of generalized items as entries. The generalized item difficulties can then be summarized by calculating the means across the rows. The rater severities, item difficulties, and threshold values can be estimated by taking average values of these generalized item difficulties across items, raters, and thresholds, respectively.

Table 5
Ratio (D) matrix for six generalized items

G-Items	1	2	3	4	5	6
1	1	4/149	14/24	3/177	16/97	3/355
2	149/4	1	103/6	30/64	107/47	20/145
3	24/14	6/103	1	4/125	16/70	3/258
4	177/3	64/30	125/4	1	118/28	20/94
5	97/16	47/107	70/16	28/118	1	4/226
6	355/3	145/20	258/3	94/20	226/4	1

Table 6
Logit matrix for six generalized items

G-Items	1	2	3	4	5	6
1	0	-3.62	-0.54	-4.08	-1.80	-4.77
2	3.62	0	2.84	-0.76	0.82	-1.98
3	0.54	-2.84	0	-3.44	-1.48	-4.45
4	4.08	0.76	3.44	0	1.44	-1.55
5	1.80	-0.82	1.48	-1.44	0	-4.03
6	4.77	1.98	4.45	1.55	4.03	0

In order to validate the algorithm implemented in R, we compared our analyses with the results from *Facets* using the full data set from Garner and Engelhard (2009). Table 7 shows the unstandardized and standardized estimates of the rater, item, and threshold parameters from both programs. The differences are quite small which indicate good conformity between *Erma* and *Facets*.

Table 7
Comparison of Erma and Facets results using complete data (Garner & Engelhard, 2009)

		Unstandardized		Standardized		
		Erma	Facets	Erma	Facets	Differences
Rater	1	-1.13	-0.83	-0.91	-0.88	-0.03
	2	-0.10	0.07	0.08	-0.08	0.16
	3	1.23	0.75	0.82	0.96	-0.13
Item	1	-0.94	-0.80	-0.87	-0.73	-0.14
	2	0.06	0.04	0.05	0.05	0.00
	3	0.88	0.75	0.82	0.69	0.14
Threshold	1	-1.88	-1.30	-1.42	-1.47	0.04
	2	1.88	1.30	1.43	1.47	-0.04
Mean		0	0	0	0	0
SD		1.28	0.91	1.00	1.00	0.11

In summary, pairwise algorithms provide elegant and effective approaches for teaching the basic principles of Rasch measurement theory. By simply counting the number of comparisons between observations, an adjacency matrix can be formed that can be used to obtain parameter estimates for raters, items and thresholds. The use of freeware in the form of R syntax is ideal for instructional purposes. Our preliminary analyses show a close correspondence between the estimates obtained with *Erma* and the *Facets* computer programs. We are currently expanding our new R program (Wang & Engelhard, 2014) to include model-data fit indices and several graphical features including variable maps, person, rater, item response functions, and category characteristic functions. We are also exploring aspects of graph theory to examine connectivity within rater-mediated designs based on the observation that the comparison matrix can be viewed as an adjacency matrix.

Jue Wang & George Engelhard, Jr.
The University of Georgia

References

Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573.

Choppin, B. (1968). Item banking using sample free calibration. *Nature*. 219, 870 - 872;

Choppin, B. (1985). A fully conditional estimation procedure for Rasch model parameters. *Evaluation in Education, 9*, 29–42.

Choppin, B. (1987). The Rasch model for item analysis. In D. I. McArthur (Ed.), *Alternative approaches to the assessment of achievement* (pp. 99–127). Norwell, MA: Kluwer.

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York: Routledge.

Garner, M., & Engelhard, G. (2002). An eigenvector method for estimating item parameters of the dichotomous and polytomous Rasch models. *Journal of Applied Measurement, 3*(2), 107–128.

Garner, M., & Engelhard, G. (2009). Using paired comparison matrices to estimate parameters of the partial credit Rasch measurement model for rater-mediated assessments. *Journal of Applied Measurement, 10*(1), 30–41.

Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago: MESA Press.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Education Research. (Reprinted 1980. Chicago: University of Chicago Press).

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy, 14,* 58–94.

R Core Team, (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Wang, J., & Engelhard, G., (2014). *Everyone's Rasch Measurement Analyzer (Erma): An R Package for Rasch Measurement Models*. Paper presented at the International Objective Measurement Workshop in Philadelphia, PA.

Almost the Tarde Model? Do psychology and the social sciences demand more fully developed and rigorous quantitative theories and methods than the natural sciences?

The work of Gabriel Tarde (1843-1904) has recently been republished and celebrated for its emphasis on themes of current interest in the philosophy, history, and social studies of science (Barry and Thrift, 2007; Candea, 2010; Clark, 1969; Joyce, 2002; Tarde, 1895, 1899, 1902, 1903). Tarde's focus on social existence as structured by ubiquitous network relations finds particular resonance with Actor-Network Theory (ANT; Latour, 2002, 2009; Latour and Lepinay, 2010). Tarde conceived of measurement in economics at the level of the individual, as did Rasch, but went beyond Rasch in realizing the importance and power of a nonCartesian and fundamentally social relation of subjects and objects.

Tarde's concept of the monad focuses on (a) the recovery of information at the individual level and (b) the structure common across aggregates within a measured construct. Conceptual differences from Rasch involve the latter's lack of appreciation for the roles of networked metrological traceability and social connectivity in creating a culture of second-nature intuitions informed by measurement. Though these latter are implied by item banking, instrument equating, and adaptive administration, the significance of common currencies for the exchange of economic and scientific value in human, social, and natural capital markets has not been recognized in the same way that it has been in manufactured and liquid capital markets.

Tarde was the pre-eminent sociologist of his day, though his work was pushed into the background as his student Durkheim's methods and theories came to the fore. Tarde's sense of the social involved aggregate group-level projections of resonating individual effects, such that the whole is inherently more than the sum of any given set of parts. This perspective was then eclipsed by Durkheim's sense of the sociological as the mere statistical sum of individual effects.

Though his reputation declined as time passed, in the 1920s and 1930s, Tarde's work strongly influenced the Chicago School of sociology's focus on social ecology and networked associations, which adopted his ideas around imitation as a mechanism of learning and social association. With the republication of Tarde's

Monadologie et Sociologie in the 1990s, his work has taken on new stature as a precursor of recent work in philosophy (Deleuze and Guttari) and in the social studies of science (Latour, 2002, 2009).

Tarde's sense of a monad seems to be akin to the abstract meaning of a unit of measurement, or a difference between two measures, defined by a construct theory and the empirical responses to specific questions. Latour (2009) presents Tarde as having integrated qualitative and quantitative concerns across the ontological divide separating the statistically-focused social sciences (supposedly modeled after the natural sciences) from those focusing on thick descriptions (hermeneutics, interpretation theory). Having no grasp of Rasch's contributions and the range of applications based on them, and there being no tradition actively following through on Tarde's ideas, Latour's efforts toward developing a framework of shared quantitative and qualitative ontological commitments remain incomplete (but see Fisher and Stenner (2011) for another perspective).

Tarde's critique of statistical social sciences is quite apt. I especially like Tarde's sense of the social sciences as being more quantitatively demanding than the natural sciences, which he claims is the case on the basis of the necessity of dealing with individual differences within aggregate constructs in the social sciences. The near-exact duplication of physical expressions of near-concatenatable units in the natural sciences, Tarde says, allows a concrete kind of conceptualization of units not available in the study of behavior, cognition, and social relationships. Hence follows his contention that quantification in psychology and the social sciences will mark a new level of intellectual achievement unequaled in the natural sciences.

Tarde's perspective will not, however, withstand the similarity that holds between Rasch's individual-level stochastic models and the physical phenomenon of stochastic resonance, or noise-induced order (Fisher, 1992, 2011; Gammaitoni, Hanggi, Jung, and Marchesoni, 1998; Repperger and Farris, 2010). At this level of organization, even in physics individual elements are not all so similar as to have negligible differences. Their differences in fact turn out to be essential to the process of quantification. Tarde's point is borne out, however, in that the inner structure of quantity is specified far more generally and precisely when stochastically varying individuals are taken into account than when they can be safely ignored.

And beyond these issues of measurement lie Tarde's emphases on even more important matters concerning the social nature of the forms of collective intelligence exhibited in linguistic, scientific, financial, and economic markets. Though the natural sciences may learn unanticipated lessons from psychology and the social sciences about modeling stochastic forms of invariance (Wilson, 2013), psychology and the social sciences have everything to learn from physics and chemistry as to how previously unimagined degrees of productivity might follow from the deployment of common languages and instruments traceable to reference standard, universally uniform units of measurement (Fisher, 2009).

William P. Fisher, Jr. – University of California - Berkeley

References

Barry, A., & Thrift, N. (2007). Gabriel Tarde: Imitation, invention and economy [introduction to a special issue on G. Tarde]. *Economy and Society, 36*(4), 509-525.

Candea, M. (2010). *The social after Gabriel Tarde: Debates and assessments.* London: Routledge.

Clark, T. (1969). *Gabriel Tarde: On Communication and Social Influence.* Chicago: University of Chicago Press.

Fisher, W. P., Jr. (1992). Stochastic resonance and Rasch measurement. *Rasch Measurement Transactions, 5*(4), 186-187 [http://www.rasch.org/rmt/rmt54k.htm].

Fisher, W. P., Jr. (2009). Invariance and traceability for measures of human, social, and natural capital: Theory and application. *Measurement, 42*(9), 1278-1287.

Fisher, W. P., Jr. (2011). Stochastic and historical resonances of the unit in physics and psychometrics. *Measurement: Interdisciplinary Research & Perspectives, 9*, 46-50.

Fisher, W. P., Jr., & Stenner, A. J. (2011). Integrating qualitative and quantitative research approaches via the phenomenological method. *International Journal of Multiple Research Approaches, 5*(1), 89-103.

Gammaitoni, L., Hanggi, P., Jung, P., & Marchesoni, F. (1998, January). Stochastic resonance. *Reviews of Modern Physics, 70*, 223-288.

Joyce, P. (Ed.). (2002). *The social in question: New bearings.* London: Routledge.

Latour, B. (2002). Gabriel Tarde and the end of the social. In *The social in question: New bearings* (pp. 117-132). London: Routledge.

Latour, B. (2010). Tarde's idea of quantification. In M. Candea (Ed.), *The social after Gabriel Tarde: Debates and assessments* (pp. 145-162). London: Routledge.

Latour, B., & Lepinay, V. A. (2010). *The science of passionate interests: An introduction to Gabriel Tarde's economic anthropology.* Prickly Paradigm Press.

Repperger, D. W., & Farris, K. A. (2010). Stochastic resonance—a nonlinear control theory interpretation. *International Journal of Systems Science, 41*(7), 897-907.

Tarde, G. (1895/2012). *Monadology and sociology* (T. Lorenc, Trans.). Melbourne, Australia: Re.press.

Tarde, G. (1899/2009). *Social laws: An outline of sociology* (H. C. Warren, Trans.). Kitchener, Ontario, Canada: Batoche Books.

Tarde, G. (1902/2007). Economic psychology (A. Toscano, Trans.). *Economy and Society, 36,* 614-643. (Reprinted from *Psychologie Economique* (pp. 77-116, 154-181). Paris: F. Alcan.)

Tarde, G. (1903). *The laws of imitation* (E. C. Parsons, Trans.). New York: Henry Holt and Company.

Wilson, M. (2013). Using the concept of a measurement system to characterize measurement models used in psychometrics. *Measurement, 46,* 3766-3774.

Worlds Standards Day Competition

Papers for this year's World Standards Day (October 14) competition are due August 8. First, second, and third place awards include checks of $2,500, $1,000, and $500, respectively. This year's theme is Standards Level the Playing Field. Sponsors include NIST and the Society for Standards Professionals. A Rasch-oriented paper won third prize in the 2011 competition. For more information, see http://www.ses-standards.org/?130.

4th Annual ORVOMS Summary

The fourth annual Ohio River Valley Objective Measurement Seminar (ORVOMS) was held on May 2, 2014 at the Vernon Manor Building of Cincinnati Children's Hospital Medical Center in Cincinnati, Ohio. George Karabatsos of the University of Illinois – Chicago served as the keynote speaker discussing Bayesian nonparametric Rasch modeling and providing a demonstration of his free Bayesian Regression software. Additional topics included the dichotomous Rasch model, Many-Facet Rasch model, scale construction, paired comparisons, and logistic regression with Rasch models. Presenters discussed a diversity of topics such as knee pain, social justice, performance based assessments, quantitative methods, primary care, and standards-based report cards for students. ORVOMS was attended by approximately 30 people from Kentucky, Ohio, Iowa, Minnesota, Illinois, Michigan, and North Carolina. A special thanks to Rick Ittenbach and Cincinnati Children's Hospital Medical Center for hosting.

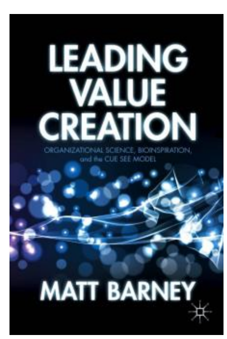
Michael Peabody
American Board of Family Medicine

New Rasch Books

Leading Value Creation: Organizational Science, Bioinspiration, and the Cue See Model.

Every business discipline has a unique vantage point on value creation and destruction, and while specialists have devised solutions, leaders rarely use them because of the inherent complexity in trying to understand which parts fit together to help them achieve goals. The result is a sort of business 'Tower of Babel' for practicing leaders and organizational scientists alike.

Leading Value Creation fills this void as the first book to take organizational science and place it into one coherent and useful model, using the latest methods in measurement. Barney integrates vastly different areas of organizational science into his Cue See Model, which builds upon his experience developing global leaders at companies like Motorola, Merck, and Infosys. The model is a way to help leaders better create value and mitigate risk. It highlights the flow of value across four perspectives—quality, cost, quantity, and cycle time, and also looks across levels of analysis for a holistic view on the bottlenecks to value creation as the best focal point for organizations to succeed. Barney provides numerous practical examples from pharmaceuticals to barbershops, and summarizes six empirical studies demonstrating the model's usefulness.



Matt Barney (2013). *Leading Value Creation: Organizational Science, Bioinspiration, and the Cue See Model*. Palgrave MacMillan. 304 pages. US$50.00. ISBN: 978-1-137-37371-7

Rasch Analysis in the Human Sciences

Rasch Analysis in the Human Sciences helps individuals, both students and researchers, master the key concepts and resources needed to use Rasch techniques for analyzing data from assessments to measure variables such as abilities, attitudes, and personality traits. Upon completion of the text, readers will be able to confidently evaluate the strengths and weakness of existing instrumentation, compute linear person measures and item measures, interpret Wright Maps, utilize Rasch software, and understand what it means to measure in the Human Sciences.

Each of the 24 chapters presents a key concept using a mix of theory and application of user-friendly Rasch software. Chapters also include a beginning and ending dialogue between two typical researchers learning Rasch, "Formative Assessment Check Points," sample data files, an extensive set of application activities with answers, a one paragraph sample research article text integrating the chapter topic, quick-tips, and suggested readings.

Rasch Analysis in the Human Sciences will be an essential resource for anyone wishing to begin, or expand, their learning of Rasch measurement techniques, be it in the Health Sciences, Market Research, Education, or Psychology.

William J. Boone, John R. Staver, & Melissa S. Yale (2014). *Rasch Analysis in the Human Sciences*. Springer. 482 pages. $129.00. ISBN: 978-94-007-6857-4

Rasch-related Coming Events

July 4-Aug. 1, 2014, Fri.-Fri. Online workshop: Practical Rasch Measurement – Further Topics (E. Smith, Winsteps), www.statistics.com

July 25, 2014, Fri. In-person workshop: Measuring Rehabilitation Outcomes in Older Adults, Chicago, www.rehabmeasures.org

July 28-Nov. 22, 2014, Mon.-Sat. Online course: Introduction to Rasch Measurement Theory (D. Andrich, I. Marais) www.education.uwa.edu.au/ppl/courses

Aug. 2-6, 2014, Sat.-Wed. PROMS2014, Guangzhou, China: Sat.-Sun. workshops; Mon.-Wed. symposium, www.confchina.com

Aug. 8-Sept. 5, 2014, Fri.-Fri. Online workshop: Many-Facet Rasch Measurement (E. Smith, Facets), www.statistics.com

Sept. 3-5, 2014, Wed.-Fri. IMEKO International Measurement Confederation Symposium, Madeira Island, Portugal, www.imekotc7-2014.pt

Sept. 10-12, 2014, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Sept. 12-Oct. 24, 2014, Fri.-Fri. Online workshop: Rasch Applications, Part I: How to Construct a Rasch Scale (W.P. Fisher), www.statistics.com

Sept. 15-17, 2014, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK

Sept. 18-19, 2014, Wed.-Fri. In-person workshop: Advanced Rasch (A. Tennant, RUMM), Leeds, UK

Sept. 30, 2014, Tues. Submission deadline: 6th Rasch Conference: Sixth International Conference on Probabilistic Models for Measurement in Education, Psychology, Social Science and Health, Cape Town, South Africa, www.rasch.co.za/conference.php

Call for Submissions

Research notes, news, commentaries, tutorials and other submissions in line with *RMT*'s mission are welcome for publication consideration. All submissions need to be short and concise (approximately 400 words with a table, or 500 words without a table or graphic). The next issue of *RMT* is targeted for September 1, 2014, so please make your submission by August 1, 2014 for full consideration. Please email Editor\at/Rasch.org with your submissions and/or ideas for future content.