Summer 2016, VOL. 30, NO. 1; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

Research Notes and Editorials Include:

Transactions of the Rasch Measurement SIG
American Educational Research Association

Polytomous Rasch Models Derived from Objectivity

Objectivity is the requirement that the measures produced by a measurement model be sample-free for the agents (test items) and test-free for the objects (people). Sample-free measurement means “item difficulty estimates are as independent as is statistically possible of whichever persons from the same population, and whatever distribution of person abilities, happen to be included in the sample.” Test-free measurement means “person ability estimates are as independent as is statistically possible of whichever items from the same population, and whatever distribution of item difficulties, happen to be included in the test.” In particular, the familiar statistical assumption of a normal (or any known) distribution of model parameters or empirical data is not required.

This also implies that Rasch estimates are statistically invariant when the data fit the Rasch model. “The argument for invariance may be stated rather loosely as follows. Irrelevancies in the data should not make a fundamental difference in the results obtained from the analysis of the data.” *(International Encyclopedia of Statistics, art. Estimation: point estimation).* For Rasch measurement, irrelevancies include the person and item distributions.

Objective measures from paired observations

Paralleling a derivation of the Rasch dichotomous model[1], this derivation is based on the hypothetical administration of numerous independent replications of the same item, *i*, to two people, *m* and *n*. The items share the same rating scale with ordered categories number 0,1,2,..,k-1,k,..,K. After L such administrations, we have the following contingency table for responses in higher category, *k*, and lower adjacent category, *k-1*, of the rating-scale, when *m* and *n* both respond in those two categories of the same item:

where Cmiknik is the count of times when both persons *m* and *n* respond in category *k* of the same administration of item *i*. This is Cmi(k-1)ni(k-1) when they respond in category *k-1*, and similarly for the other cells.

Let us compare the performances of persons on categories *k-1* and *k*. In those instances when both *m* and *n* answer in the same category in the same item administration, we detect no difference in their performances. Consequently the informative contrast of their performance is the comparison between **Cmi(k-1)nik** and **Cmikni(k-1)**. The ratio of these terms, **Cmi(k-1)nik / Cmikni(k-1)** is a comparison of the frequencies of the responses of the two people on the L administrations of item *i*. In the limit as L becomes larger, the frequencies become their probabilities multiplied by the number of replications, L. Thus, since the responses by persons *m* and *n* are independent,

$$\text{Cnikmik} = \text{Prob}(n,i,k) * \text{Prob}(m,i,k) * L$$

where Prob(n,i,k) is the probability that person *n* responds in category *k* of item *i*, and similarly for the counts in the other cells. Thus

$$\text{Cnikmi}(k-1) = \text{Prob}(n,i,k) * \text{Prob}(m,i,k-1) * L$$

$$\text{Cni}(k-1)\text{mik} = \text{Prob}(n,i,k-1) * \text{Prob}(m,i,k) * L$$

so that

$$\frac{\text{Cnikmi}(k-1)}{\text{Cni}(k-1)\text{mik}} = \frac{\text{Prob}(n,i,k) * \text{Prob}(m,i,k-1)}{\text{Prob}(n,i,k-1) * \text{Prob}(m,i,k)}$$

Use of Objectivity

What happens when we require this comparison to maintain objectivity? Then the comparison of the performance of persons *n* and *m* must not depend on which particular item we use to compare them. If we choose to use item *j* we must obtain the same result. Expressing this algebraically:

$$\frac{\text{Prob}(n,i,k) * \text{Prob}(m,i,k-1)}{\text{Prob}(n,i,k-1) * \text{Prob}(m,i,k)} = \frac{\text{Prob}(n,j,k) * \text{Prob}(m,j,k-1)}{\text{Prob}(n,j,k-1) * \text{Prob}(m,j,k)}$$

Rearranging,

$$\frac{\mathrm{Prob}(n,i,k)}{\mathrm{Prob}(n,i,k-1)} = \frac{\mathrm{Prob}(n,j,k)}{\mathrm{Prob}(n,j,k-1)} * \frac{\mathrm{Prob}(m,i,k)}{\mathrm{Prob}(m,i,k-1)} * \frac{\mathrm{Prob}(m,j,k-1)}{\mathrm{Prob}(m,j,k)}$$

rating-scale categories, *k*. It is related to early forms of the polytomous Rasch model[2].

Counts of Responses in Pairs of Categories after L Administrations of Item i									
	Person *n*								
	Category	**0**	**1**	**...**	**k-1**	**k**	**...**	**K**
Person *m*	**0**	Cmi0ni0	Cmi0ni1	..	Cmi0ni(k-1)	Cmi0nik	...	Cmi0niK
	1	Cmi1ni0	Cmi1ni1	...	Cmi1ni(k-1)	Cmi1nik	...	Cmi1niK
	...	....	...	...	...	...	...	...
	k-1	Cmi(k-1)ni0	Cmi(k-1)ni1	..	Cmi(k-1)ni(k-1)	**Cmi(k-1)nik**	...	Cmi(k-1)niK
	k	Cmikni0	Cmikni1	...	**Cmikni(k-1)**	Cmiknik	...	CmikniK
	...	...	...	...	...	..	...	...
	K	CmiKni0	CmiKni1	...	CmiKni(k-1)	CmiKnik	...	CmiKniK

However, again by objectivity, the interaction of person *n* and item *i* must not depend on which person *m* and which item *j* is used for comparison in the measuring process. Consequently we can choose the measure of person 0 to define the frame of reference for the persons and the measure of item 0 to define the frame of reference for the items and so provide fixed reference points on the person and item measurement scales. Thus

$$\frac{\mathrm{Prob}(n,i,k)}{\mathrm{Prob}(n,i,k-1)} = \frac{\mathrm{Prob}(n,0,k)}{\mathrm{Prob}(n,0,k-1)} * \frac{\mathrm{Prob}(0,i,k)}{\mathrm{Prob}(0,i,k-1)} * \frac{\mathrm{Prob}(0,0,k-1)}{\mathrm{Prob}(0,0,k)}$$

We see that Prob(n,0,k)/ Prob(n,0,k-1) is independent of item i, and depends only on the relationship of person *n* to the reference person and item, and to the choice of the pair of categories indicated by *k*. Let us call this term, b_{nk}. Similarly, e_{ik} for item i, and f_k for the reference item and person.

$$\frac{\mathrm{Prob}(n,i,k)}{\mathrm{Prob}(n,i,k-1)} = b_{nk} * e_{ik} / f_k$$

Taking logarithms,

$$\log\left(\mathrm{Prob}(n,i,k) / \mathrm{Prob}(n,i,k-1) \right) = B_{nk} - D_{ik} - F_k$$

where $B_{nk} = \log(b_{nk})$, $D_{ik} = -\log(e_{ik})$, $F_k = \log(f_k)$

This general form has different abilities for person *n* and different item difficulties for item *i* depending on the choice of the adjacent pair of rating-scale categories, *k*.

However, let us make this model yet more objective by constraining the values of B_{nk} to be independent of the pair of categories denoted by subscript *k*, so that $B_{nk} = B_n$, and similarly $D_{ik} = D_i$, then we have the polytomous Rasch model known as the Andrich Rating Scale Model[3] (RSM) written in logistic form:

$$\log\left(\mathrm{Prob}(n,i,k) / \mathrm{Prob}(n,i,k-1) \right) = B_n - D_i - F_k \quad \textit{for } k=1 \textit{ to } K$$

where B_n is defined as the ability of person *n;* D_i is the difficulty of item *i;* F_k is the Andrich threshold, indicating the point on the latent variable at which categories *k-1* and *k* are equally probable. This logistic form of the RSM model can be rewritten in the familiar exponential form. A similar derivation produces the Partial Credit Model.

J. M. Linacre and B. D. Wright

[1] "Dichotomous Rasch model derived from objectivity." Wright BD, Linacre JM (1987). *Rasch Measurement Transactions*, 1(1), 5-6 www.rasch.org/rmt/rmt11a.htm

[2] "Some remarks concerning inference about items with more than two categories." Georg Rasch, Jon Stene (1967) www.rasch.org/memo/memo1967.pdf

[3] "A rating formulation for ordered response categories." Andrich, D. (1978). *Psychometrika, 43*, 561-73.

Inference of Independent Dichotomous Responses in the Polytomous Rasch Model

Arising from a presentation on the history of Georg Rasch's and Ben Wright's struggles with a unidimensional model with sufficient statistics for responses in more than two ordered categories, as shown by their correspondence on the topic (Andrich, 2016), a question which arose was: could *independent* dichotomous responses at the thresholds be reconstructed from a polytomous response in the ordered categories? This note is an answer to that question.

The answer is that even if it were possible to do that, it is unnecessary. It is unnecessary because the analysis of a polytomous response to an item using the polytomous Rasch model (PRM) gives estimates of threshold and person parameter estimates *as if* the responses at the thresholds were independent dichotomous ones and the responses were analysed using the dichotomous Rasch model. This is true at the item level whether the parameterisation across the items is rating scale or partial credit (Wright and Masters, 1982). It is one of a number of remarkable properties of the PRM. Below is a brief explication of this property.

To be concrete with a response in three putatively ordered categories, suppose that they are as shown in Figure 1 in which narrative essays were classified according to the quality of *setting* characterised by the descriptors in the second row of the table. Although the response in one of the categories is taken as probabilistic, once the response is obtained, it is set in one of the three categories intended to be ordered. The structure of the ordered categories is independent of the way the response is generated. Figure 1 also shows the boundaries, the thresholds, on the continuum between the categories. The second threshold is simply an extension of having one threshold on the continuum in the case of a dichotomous response.

The first implication of the categories being ordered is that a transitive property on the continuum must hold. Thus, if the response is in the category *Inadequate*, the implication is that the essay was not only considered worse than *Adequate*, but was also considered worse than *More than adequate*. That is, the essay fails at both the thresholds. Expressing this transitivity as a failed dichotomous response (0) at both thresholds *resolves* the single response to the implied pair of dichotomous responses (0,0), where the ordered pair signifies responses at the successive thresholds. Similarly, if the response was *Adequate*, the implication is that the essay was considered better than *Inadequate*, but simultaneously not as good as *More than adequate*. Expressing this transitivity as a success (1) at the first threshold and a failure at the second threshold, resolves the single response to the implied pair (1,0). Finally, if the response is *More than adequate*, it implies that the essay was considered not only better than *Adequate*, but also better than *Inadequate*. Resolving this response into the two implied dichotomous responses gives (1,1).

Table 1 shows this set of implied dichotomous responses, together with the expression for the PRM for each of the responses for person n with location parameter β_n and threshold difficulties δ_{i1}, δ_{i2} of item i. The PRM is derived from assuming the dichotomous Rasch model at each threshold, but constraining the probabilities to the patterns shown in Table 1 so that they sum to 1.0.

We note now that in Table 2, it was not possible to construct the response (0,1) which implies failing the first (*Inadequate*) but *simultaneously* passing the second threshold (*More than adequate*). Such a response contradicts the constraint of order on the continuum. The set of responses shown in Table 1 is known as a Guttman pattern (Guttman, 1950).

The remarkable property of the PRM considered here is that the parameters estimated from analysis of data that fit the PRM are the *same* parameters as in an analysis of dichotomous responses with the dichotomous Rasch model that include the response (0,1). The set of Guttman patterns shown in Table 1 is taken as a *subset* of responses of the full set of locally independent dichotomous responses at the thresholds which fit

Inadequate	Adequate	More than adequate
None	Discrete	Integrated and manipulated

δ_{i1} δ_{i2} $\rightarrow$

Figure 1: Classification of essays into one of three ordered categories with respect to setting

the dichotomous Rasch model with the same threshold values. In a full set of locally independent responses, the pattern (0,1) would appear occasionally. The PRM accounts for the structural dependence amongst the responses in the Guttman pattern which reflect the constraint of order in a single response. One manifestation of this structural dependence is that the probability of a response in the first category is a function of the values of both thresholds, not just the first one, as is evident from the denominator of Table 1.

To illustrate this relationship concretely with an example, I simulated the responses of 5000 people to four locally independent dichotomous items according to the dichotomous Rasch model. These four items were considered below in two sets of two items each.

The first analysis involved all four items analysed simultaneously with the dichotomous Rasch model. For the second analysis with the PRM, the following steps were taken. First, for each set of items, the subset of dichotomous responses which formed a Guttman pattern according to the hypothesised ordering of their difficulty was selected, and the non-Guttman patterns ignored. Second, for those responses which conformed to the Guttman pattern in each set, the scores of dichotomous items were summed, giving a structure as in Table 1. Thus now there were two polytomous items each with a maximum score of two, where each score reflected the number of thresholds exceeded. The scores on these two constructed, polytomous items were analysed with the PRM.

Table 2 shows an excerpt from the data matrix. A non-Guttman pattern is treated simply as missing data. Clearly, the number of pairs of responses with a non-Guttman pattern would ideally not be large, and given the values of threshold parameters, the probability of the non-Guttman pattern can be calculated. It can be shown readily that for a given total score, the Guttman pattern has the highest probability (Andrich, 1985).

Table 3 shows the simulated parameters and their estimates using consistent, pairwise conditional estimates implemented in RUMM2030 (Andrich, et. al. 2015) from the two analyses. It also shows the number of persons in each analysis. Because of extreme scores, the number in the first analysis used in the estimation is less than the 5000 simulated, and because of both extreme scores and the use of only Guttman patterns, the number in the second analysis is even smaller. It is evident that the estimates in both cases are excellent, with the estimates of the first pair of items slightly better in the dichotomous model analysis, and the estimates of the second pair slightly better in the PRM analysis. The estimates are not identical because the data are not identical, the data in PRM anlaysis being a subset of the dichotomous model analysis. However, the two analyses estimate identical parameters.

To stress this identity between the analysis with the dichotomous Rasch model and the PRM, the first panel of Figure 2 shows the dichotomous item characteristic curves (ICCs) for the two items of Set 1 and the second panel shows the identical latent threshold characteristic curves (TCC) for the corresponding polytomous item. The latent curves in the second panel are shown as dotted lines because in a response in ordered categories, these dichotomous responses at the thresholds are latent, they are never observed.

Table 1. Resolved latent dichotomous responses at the thresholds and the PRM

Number of thresholds exceeded in the observed response	Resolved implied dichotomous responses at the thresholds	The PRM
0	(0,0)	$1/\gamma_{ni}$
1	(1,0)	$\exp(1\beta_n - \delta_{i1})/\gamma_{ni}$
2	(1,1)	$\exp(2\beta_n - \delta_{i1} - \delta_{i2})/\gamma_{ni}$
$\gamma_{ni} = 1 + \exp(\beta_n - \delta_{i1}) + \exp(2\beta_n - \delta_{i1} - \delta_{i2})$		

Table 2. Item parameters for the simulation and an excerpt of the data structure for two analyses

Dichotomous Items				Guttman Subset	
Set 1		Set 2		Set 1	Set 2
Item 1	Item 2	Item 1	Item 2	Item 1	Item 2
1	0	0	1	1	
1	0	1	1	1	2
0	0	1	0	0	1
0	1	1	0		1
.	.	.	.	.	.
.	.	.	.	.	.
1	1	1	0	2	1

Table 3. Simulated item parameters and estimates from two analyses

	Analysis with the dichotomous model N=4008				Analysis with the PRM N = 3436			
	Dichotomous items Set 1		Dichotomous items Set 2		Polytomous thresholds Set 1		Polytomous thresholds Set 2	
	δ_{11}	δ_{12}	δ_{21}	δ_{22}	δ_{11}	δ_{12}	δ_{21}	δ_{22}
Simulated	-0.75	0.75	-1.00	1.00				
Estimate	-0.75	0.76	-1.04	1.03	-0.72	0.73	-1.02	1.01

$\beta \sim N(0,1)$

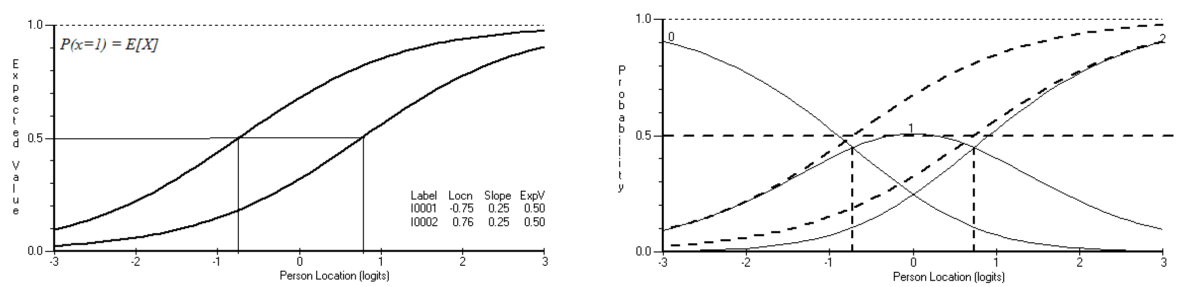

Figure 2 Dichotomous ICCs of items of Set 1 (Panel 1) and (dotted) latent dichotomous TCCs of Set 1 (Panel 2).

In summary, the analysis of a polytomous item with the PRM is *as if* the responses were a Guttman subset of the full set of locally independent dichotomous responses at the thresholds between the adjacent categories which are analysed with the dichotomous Rasch model. All interpretations of the threshold estimates are compatible with this relationship. Thus in answer to the original question: there is no need to try to construct dichotomous responses from the polytomous response – the analysis is identical to an analysis *as if* the responses at the thresholds were independent dichotomous responses analysed with the dichotomous Rasch model. Derivations of this relationship between the dichotomous Rasch model and the PRM can be found in Andrich (1978, 2010), where two other remarkable properties of the PRM are explained, first that the frequencies in two adjacent categories can only be combined if the discrimination at the threshold between them is 0; second, that the estimates of the thresholds treats the empirical ordering of the categories as an hypothesis rather than taking it for granted, and that as when the discrimination between the thresholds is 0 or there are other problems with the responses, the hypothesis may be rejected by a lack of ordering of the estimates of the thresholds.

David Andrich, *University of Western Australia*

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43* (4), 561-574.

Andrich, D. (l985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological Methodology*, San Francisco, Jossey-Bass. (Chapter 2, pp. 33-80.).

Andrich, D. (2010) Understanding the response structure and process in the polytomous Rasch model. In *Handbook of Polytomous Item Response Theory Models: Developments and Applications*. M. Nering and R. Ostini (Eds.) Lawrence Erlbaum Associates, Inc. Chapter 6: pp 123 – 152.

Andrich, D. (2016) Georg Rasch and Benjamin Wright's struggle with the unidimensional polytomous model with sufficient statistics. *Educational and Psychological Measurement*. Published ahead of print: http://epm.sagepub.com/content/early/2016/02/29/0013164416634790.full.pdf+html.

Andrich, D., Sheridan, B.S. and Luo, G. (2015) RUMM2030: Rasch Unidimensional Models for Measurement. RUMM Laboratory: Perth, Western Australia. www.RUMMLab.com.

Guttman, L. (1950). The basis for scalogram analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star and J.A. Clausen (Eds.), *Measurement and Prediction*, pp.60-90. New York: Wiley.

Wright, B.D. & Masters, G.N. (1982). *Rating Scale Analysis: Rasch Measurement.* Chicago: MESA Press.

Rasch Measurement Theory and Applications Conference

The University of Western Australia is pleased to announce the Seventh International Conference on Probabilistic Models for Measurement, to be held at The University Club on the Matilda Bay of the Swan of River. The conference will cover the range of areas where Rasch measurement theory is applied: education, psychology, health, marketing and social science. Full details of fees and processes for submission of abstracts will be available in June 2017 at: http://www.education.uwa.edu.au/ppl/raschconference/.

Also of note, preceding the conference there will be a five day *Advanced Course in Rasch Measurement Theory and the Application of RUMM2030* from Wednesday 10 January 2018 to Tuesday 16 January 2018. Please see (http://www.education.uwa.edu.au/ppl/raschconference/course) for more information about this course.

Review of IOMW 2016

IOMW 2016, the western companion conference to PROMS, was held April 4-6 in Washington, DC. Hosted by the lovely Mount Vernon United Methodist Church near the center of the city, ringed with cherry blossoms, the conference was by all accounts a brilliant success and a solid step forward in the work of the international measurement community.

The unusually well-attended pre-conference on Monday, April 4, was dedicated to software. Eight software packages were presented -- four old standards and four relatively new packages: RUMM (David Andrich), Winsteps (William Boone), Facets (Mike Linacre), ConQuest (Rebecca Freund), ERMA (George Engelhard, Jr.), OpenBUGS (Hong Jiao), jMetrik (Patrick Meyer), and Damon-on-Python (Mark Moulton). After the presentations, there was a 2-hour panel on "the status of measurement software" and future directions. Moulton concluded by recommending the "scientific python" community (SciPy) as an exemplar for building an open-source library around a common language and set of tools for the measurement community. He called the proposed library MOMS (Multidisciplinary Open Measurement Software). It would consist of a "big data" utility, a cross-discipline library of psychometric and other measurement functions, and tools for running community software such as R, Winsteps, and jMetrik from inside Python. The day closed with workshops and round-table one-on-ones.

On Tuesday, Mark Wilson raised the very real prospect of Rasch measurement models being absorbed into statistics libraries as just another analysis method. He proposed emphasizing measurement as a human-driven interface between content knowledge and statistics, an approach being pursued through the Bear Assessment System Software (BASS) project. Other "foundational" papers addressed validity (Duckor, Behizadeh), the mathematical foundations of special objectivity (San Martin, Avello), and Rasch models as way to model supply and demand and other social interactions (Fisher).

Along this line, William Fisher promoted participation by Rasch professionals in the upcoming IMEKO 2016 conference on international scientific metrology (http://imeko-tc7-berkeley-2016.org/) to be held in Berkeley August 3-5. IMEKO offers a critical opening for broadening the application of Rasch models beyond education and health outcomes.

Also on Tuesday morning were an Andrich retrospective on Georg Rasch's struggle with polytomous models, a polytomous model to guarantee ordered thresholds (Chris Bradley), and an important presentation by Mike Linacre on estimation challenges that occur when applying JMLE to sparse datasets, with ways to overcome them.

Tuesday afternoon was a blizzard of presentations dealing with analysis of fit, DIF, testlets, multidimensionality, validation, detection, classification, and test construction. The presentations were startling and innovative, both in the choice of subject matter and in the integration of new methods with Rasch ideas.

Wednesday morning opened with a session on non-cognitive measurement as used for NAEP, PISA, and college success. The college success paper involved a quality of spiritual life construct build by the Church of the Latter Day Saints. Skye Barbic shared her work on measuring recovery outcomes in psychiatry, where patients, doctors, and social workers could visualize their progress on a plastic centimeter ruler. There were presentations on measuring household food security, metrological standards for health-care decision-making, non-cognitive readiness scales, metacognition, and teacher evaluation in Chile.

Wednesday afternoon had a session on issues of comparability and stability relating to adolescent mental health (Curt Hagquist), repeated measures (Vernon Mogol), and vertical scaling (Ida Marais). There was a session on the use of rating scale models to measure writing and musical performance. Another session focused on growth, including a striking demonstration of the use of lexile and quantile growth curves in North Carolina (Gary Williamson) and a mathematical specification of a value-added model for school accountability in Chile (Veronica Santelices and Ernesto San Martin). William Fisher philosophized on the "curves of life" described by Cook in 1914 -- how it is precisely the aberrations from structure that make life possible, even as such aberrations are only possible when a clear structure already exists.

For the last session, David Andrich discussed Rasch's criterion of invariance and Jack Stenner (standing in for Mark Stone) demonstrated the practical consequences of measurement invariance and the objective specification of scales. Travel awards of $250 were then awarded to Pey Shin Ooi (University of Adelaide), Hsiu-yi Chao (National Chung Chen University), Manqian Liao (University of Maryland, College Park), Jue Wang (University of Georgia), Chi-Chen Chen (National Sun Yat-Sen University) and Jinho Kim (University of California, Berkeley). Jinho Kim also received an honorable mention for his amazingly innovative paper, "Polytomous extension of item explanatory Rasch models: an application to the carbon cycle assessment data".

Sarah Thomas (University of Virginia) received the Benjamin Wright Innovations in Measurement Award ($500) for her beautifully presented paper combining Rasch modeling with machine learning, "Identifying compromised test items using the Rasch model and support vector machines." Dandan Liao (University of Maryland, College Park) received the Best Paper by a Graduate Student award ($500) for her advanced and elegant paper, "A multigroup cross-classified testlet model for dual local item dependence in the presence of DIF items." Presentations and papers are being collected and will be posted on the iomw.org website, along with photographs and other memorabilia.

The IOMW Conference Committee, under the leadership of Brent Duckor and Mark Moulton, passed the IOMW torch to the formidable team of Ronli Diakow and Andrew Maul. Ronli and Andy have kindly consented to organize the next IOMW, to take place in New York City in April 2018.

The conference concluded with a toast and memorial for Benjamin Wright, the passionate and beloved champion of Rasch measurement in the 20th century, who passed away last October.

Mark Moulton, *IOMW Conference Committee*

Journal of Applied Measurement
Vol. 17, No. 2, 2016

Richard Smith, Editor, www.jampress.org

AERA 2016 Rasch SIG Business Meeting Update

The Rasch SIG Business Meeting took place on Monday, April 11. Despite the SIG meeting taking place a full 5 days after IOMW had wrapped up, the meeting still attracted an audience of approximately 20-25 attendees.

The session began by thanking outgoing chair Jim Tognolini and other members of the SIG leadership team for their service to the SIG. Next, results of the recent elections for the 2016-2018 term were announced, which included Leigh Harrell-Williams elected as SIG chair, Mikaela Raddatz elected SIG Secretary and Matt Schulz elected SIG Treasurer. After a briefing on SIG finances and other business news, David Andrich was presented the inaugural ***Benjamin Drake Wright Senior Scholar Award***.

(Pictured: Kenneth Royal (right) presents David Andrich (left) the Benjamin Drake Wright Senior Scholar award).

Attendees were then treated to a very engaging presentation entitled “Applying the Rasch model to assess cross-cultural comparability of test scores” from guest speaker Elena Kardanova from the National Research University Higher School of Economics in Moscow, Russia.

Per usual, the SIG meeting was a great opportunity for networking, catching up with old friends and making some new ones. We’ll look forward to seeing everyone next year in San Antonio, Texas!

Thales and Rasch

Teaching the geometry of the right triangle is frequently enhanced by the story of Thales (c. 640-550 BCE) who applied the property of similar triangles to measure the height of a pyramid. He deduced that the shadow of the pyramid (QB) and, simultaneously, that of a vertical stick (AB) placed at the end of the shadow of the pyramid produce the following property of similar triangles: The height of the pyramid (PQ) is to the height of the stick (AB), as the length of the shadow of the pyramid (QB) is to the length of the shadow of the stick (BC), i.e. PQ/AB = QB/BC. Two comparisons to Rasch are evident in this illustration:

1. The practical solution to finding the height of any object is determined by an abstract principle, not by data. Rasch always sought to find general solutions to measurement problems not merely to produce a description of data (Stone & Stenner, 2014, 2016). Similarly, we follow Newton's theory of gravitation not his data.

2. The solution to the problem of determining the height of any object is *independent* of the height of the stick utilized, and *independent* of the height of the object that is to be determined. Thus, we see that parameter separation and general objectivity are very old and very useful ideas.

Mark Stone and Jack Stenner

References

Stone, M. & Stenner, A. J. (2014). Comparison is key. *Journal of Applied Measurement*, *15*(1), 26-39.

Stone, M. & Stenner, A. J. (2016). Rasch's growth model. Under review for publication.

The Impact of “Agent” and “Patient” Roles on Items

As item writers, we do our best to follow established guidelines and “best practices” (Haladyna, Downing & Rodriguez, 2002; Nardi, 2006) for constructing items that are technically sound and error free. However, despite a strong body of research in this area, on occasion, new validity threats surface as we learn more about subconscious human biases. One example emanates from the work of psychology researchers at Yale University.

Experimental Studies

A number of researchers in the fields of cognitive science and psychology have studied the effects of *agents* and *patients*. In short, an *agent* is an individual that intentionally brings about an outcome, and a *patient* is an individual who suffers the effects of that outcome. Researchers in the field of linguistics have also studied the nature of agents and patients. Although there are a number of competing theories, experts in linguistics tend to agree that people generally interpret information based on agent and patient roles so as to help better understand events. Strickland, Fisher and Knobe (2012) describe several very interesting research studies in which they investigated how individuals respond based on the arrangement of a variety of sentences.

In their experiments, the authors constructed sentences that were logically symmetric and say or mean the exact same thing, with the only difference being the role of the agent and the patient. They used the following example to illustrate a logically symmetric item pair whose meaning should be inter-changeable:

(1) John sold products to Susan.
(2) Susan bought products from John.

The authors found “when people actually interpret these sentences they have an automatic bias to attribute more agency to grammatical subjects compared to non-subjects” (p. 200). In a second study, the researchers again tested the hypothesis that agent and patient roles affect participant responses by presenting participants with the following logically-symmetric item pair:

(1) John French-kissed Susan.
(2) Susan French-kissed John.

The authors noted they used the term “French-kiss” because it is a two-directional action, whereas a (traditional) kiss may be purely one-directional. The meaning of these two sentences represents the exact same action, therefore in theory it should not matter which individual is presented as the agent and which is presented as the patient. Again, however, results of their study indicated participants interpreted each sentence differently by consistently assigning an intentionality (agent) role to the subject of the sentence, as opposed to the object.

Implications for Measurement

So, what exactly does this mean for objective measurement? Results from Strickland et al.’s experimental studies suggest people tend to subconsciously assign a variety of roles to subjects and objects in sentences. Thus, it is entirely plausible that participants may inadvertently attach meaning to items (e.g., assumptions about intent, moral/ethical judgments, etc.) based on the presentation of the subject and object of the statement. This has significant implications for objective measurement as intent, whether it is factual and evidentiary, or opinion and interpretation, can introduce additional noise and perhaps even dimensionality. Further, the presentation of items coupled with hidden intent could cause items to change locations on the construct hierarchy (depending upon how they are presented) and affect invariance.

Practically speaking, it may be quite difficult to discern the effects of agent and patient roles (and hidden intent) on test and/or survey items without rigorous pilot testing and qualitative feedback. We often forget that language is a symbolic representation of meaning at best, and sometimes the choice of words we use communicates more than we intend. Further research is needed to better understand the impact of agent and patient roles on objective measurement.

Kenneth Royal, *North Carolina State University*
Melanie Lybarger, *Independent Consultant*

References

Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-334.

Nardi, P. (2006). *Doing survey research: A guide to quantitative methods*. Allyn & Bacon Publishers.

Strickland, B., Fisher, M., & Knobe, J. (2012). Moral structure falls out of general event structure. *Psychological Inquiry, 23*(2), 198-205.

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

Rasch Analysis for Continuous Variables

Rasch is formulated in terms of distinguishable qualitative advances, but merely applied to discretely ordered scales (Linacre, 2007). We developed an MS Excel-based module (Figure 1) which could effectively execute continuous observations besides being used for dichotomous and polytomous scores (Table 1 and 2) (Linacre, 1994; 1997).

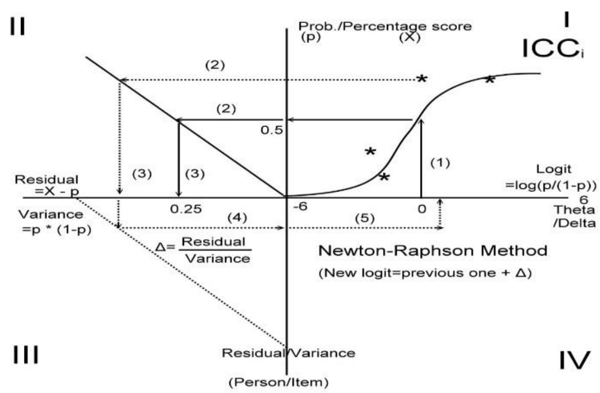

Figure 1. Process of Rasch analysis for continuous variables

We illustrated continuous variables of 16 medical reimbursement fees in a hospital for 361 inpatient cases in December 2015 within a specific major diagnosis category (MDC) of diagnosis related groups (DGRs) (Chien & Djaja, 2015). Thirty five cases were screened out due to their Outfit mean square errors greater than 2.0 (Linacre, 2002). The Wright map (Wilson, 2011) and the Kidmap diagram (Chien, Wang, Wang, & Lin, 2009; Masters, 1994) intended to be used mainly as detection tools and for obtaining clues about the possible causes of the detected misfit can be successfully plotted on MS Excel(Figure 2 and 3). The module is available for download at http://www.healthup.org.tw/continous_rasch.zip.

Table 1. Output comparison for dichotomous scores

Types of Patterns	Responses	Winsteps Measure	Winsteps Out.MNSQ	This study Measure	This study Out.MNSQ
Modelled/Ideal	1110110110100000	0.000	0.560	0.000	0.640
Guttman	1111111100000000	0.000	0.580	0.000	0.654
Miscode	0000000011111111	0.000	3.490	0.000	2.823
Carelessness	0111111110000000	0.000	0.700	0.000	0.736
Lucky Guessing	1111111000000001	0.000	0.820	0.000	0.852
Miskey	1010101010101010	0.000	1.260	0.000	1.256
Special Knowledge	1111000011110000	0.000	0.960	0.000	1.039
Imputed Outliers	1111010110010000	0.000	0.750	0.000	0.820
Low Discrimination	1110101010101000	0.000	0.730	0.000	0.809
High Discrimination	1111110101000000	0.000	0.750	0.000	0.820
Much Discrimination	1111111010000000	0.000	0.410	0.000	0.488
	Mean	0.000	1.001	0.000	0.994
	Correlation Coefficient			1.000	0.998

Table 2. Output comparison for polytomous scores

Types of Patterns	Responses	Winsteps Measure	Winsteps Out.MNSQ	This study Measure	This study Out.MNSQ
I. Model-fitting	3333313221000000101	-0.170	0.970	-0.328	1.008
	3133233232122000000	0.060	0.920	-0.002	0.966
	3333333112230000000	0.060	1.010	-0.002	1.070
	3333333111001020000	-0.110	1.030	-0.246	1.082
II. **Overfitting** (muted)	3322222222111111110	0.060	0.220	-0.002	0.183
	3333322222111110000	0.060	0.540	-0.002	0.540
	3333333322110000000	0.060	0.970	-0.002	1.017
	3222222222111111111	0.060	0.120	-0.002	0.072
	3232323212121210101	0.180	0.410	0.161	0.397
III. Limited categories	3333333333222222222	1.400	0.290	1.735	0.238
	2222222222111111111	0.000	0.060	-0.002	0.072
	3333332222222221111	0.690	0.280	0.847	0.242
IV. Informative-noisy	3222222220111111113	0.060	0.520	-0.002	0.525
	3323333221233300000	0.430	1.020	0.493	1.080
	3333333333000000000	0.060	1.370	-0.002	1.474
	3313333023230010100	0.120	1.140	0.079	1.198
V. **Non-informative**	2222222222222222222	0.560	0.210	0.756	0.223
	1212121212121212121	-0.050	0.430	-0.002	0.412

	0320200210111331100	-0.470	1.530	-0.496	1.500
	0123012301230123012	-0.110	1.460	-0.002	1.369
	0303030303030303030	-0.110	2.480	-0.002	2.495
VI. Contradictory	1111112223322211111	0.000	0.760	0.079	0.766
	2222222222333333333	1.290	1.180	1.735	1.247
	1111111111222222222	-0.050	0.990	-0.002	1.012
	0011111111222222223	-0.110	1.580	-0.002	1.464
	0000000000333333333	-0.110	4.160	-0.002	4.294
	Mean	0.148	0.987	0.184	0.998
	Correlation Coefficient			0.986	0.998

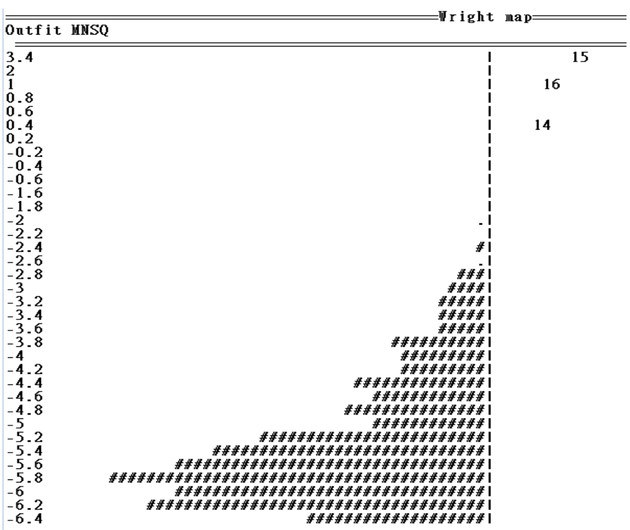

Figure 2. A Wright map plotted in the module

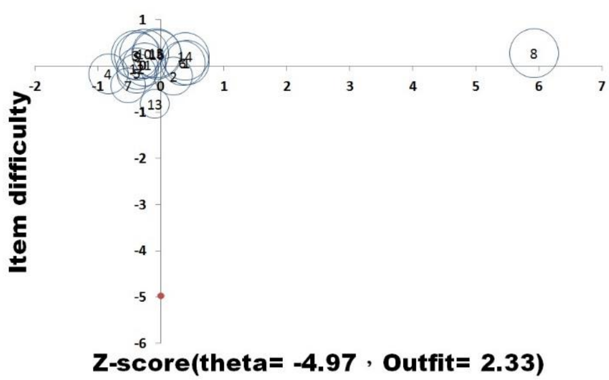

Figure 3 A Kidmap diagram plotted in the module

Tsair-Wei Chien, *Chi Mei Medical Center*
Yang Shao, *City University of Macau*

References

Chien, T. W., Wang, W. C., Wang, H. Y., & Lin, H. J. (2009). Online assessment of patients' views on hospital performances using Rasch model's KIDMAP diagram. *BMC Health Services Research*, *9*(1), 135.

Chien, T. W., & Djaja, N. (2015). Investigation and application of the person aberrant detection indices. *Rasch Measurement Transactions*, *25*(3), 1331.

Linacre, J. M. (1994). (Dichotomous Mean-square) Chi-square fit statistics. *Rasch Measurement Transactions*, *8*(2), 360.

Linacre, J. M. (1997). An all-purpose person fit statistic? *Rasch Measurement Transactions*, *11*(3), 582-583.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*(1), 85-106.

Linacre, J. M. (2007). Rasch and continuous variables. *Rasch Measurement Transactions*, *21*(1), 1088.

Masters, G. N. (1994). KIDMAP – a history. *Rasch Measurement Transactions*, *8*(2), 366.

Wilson, M. (2011). Some notes on the term: "Wright Map". *Rasch Measurement Transactions*, *25*(3), 1331.

Association of Rasch Person Fit Statistics and Latent Classes Using Social Network Analysis

The purpose of latent trait and latent class analysis (Langeheine & Rost, 1988) is to partition the sample of persons into a minimum number of homogeneous classes and to explain the data in terms of how the persons in the different classes responded differently to the items (Andrich, 1991). We rarely see any professional software, but WINMIRA (Von Davier, 2001) , to analyze person latent class using Rasch model. An approach incorporated with social network analysis (SNA) with Ucinet (Borgatti, Everett, Freeman, 2002) was proposed to show how to report person latent class in this study.

A simple dichotomous dataset (Linacre, 1994) was illustrated. After transforming the 2-mode (person in rows and item in columns) Rasch standardized residual scores into a one-mode (both person in rows and columns) metric through steps (via tools > consensus analysis > agreement), we conducted a factor analysis (via tools > scaling/decomposition) and drew a plot (via visualize > netdraw > file > open > select Loadings.##h > layout > graph theoretic layout) to show person latent classes in Figure 1.

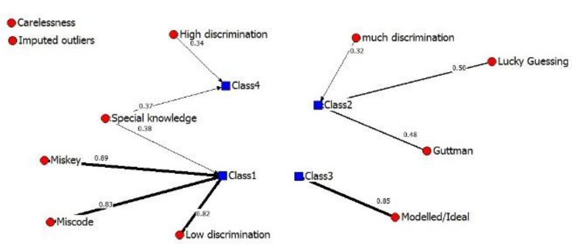

Figure 1. Using Rasch standardized residual scores to plot latent classes

Applying the non-parametric H^T fit statistic (Sijtsma, 1986; Linacre, 2012) to calculate coefficients of any paired persons as to form a one-mode metric, a slightly different plot was displayed in Figure 2. Alternatively, another simple polytomous dataset (Linacre, 1997) using point-biserial coefficients (http://www.winsteps.com/winman/correlations.htm) was applied to draw the SNA plot in Figure 3. We can see that the classes according to response patterns were also easily and separately displayed along with the momentum of Rasch fit statistics.

Rasch standardized residual scores yielded by Winsteps software or other counterparts were recommended to apply SNA for obtaining homogeneous classes and further explaining the data in terms of how the persons in the different classes responded differently to the items.

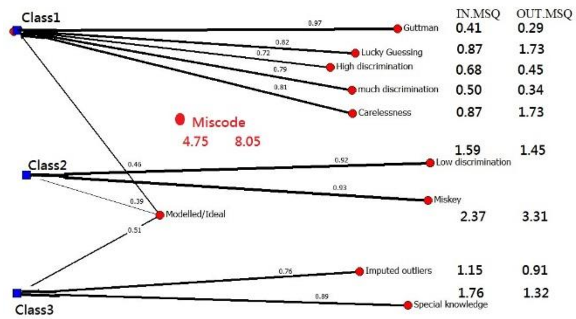

Figure 2. Using the non-parametric H^T fit statistic to plot latent classes

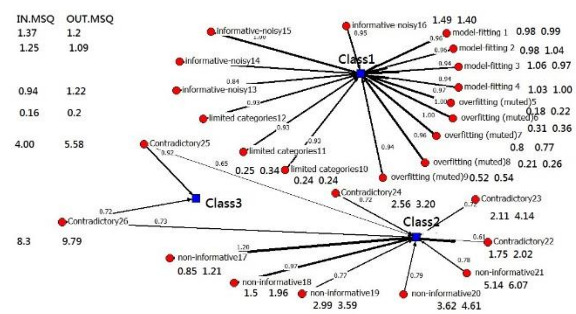

Figure 3. Using point-biserial coefficients to plot latent classes

A Youtube video is available to help readers execute Winsteps and organize data to fit the Ucinet software, and to execute several data approaches and draw the Class plot on a map. It can help us distinguish latent classes of different responses for persons, not just referring to the Rasch fit statistics. The link to the video is https://youtu.be/HF3nxaLvm0E.

Tsair-Wei Chien, Chi Mei Medical Center

References

Andrich, D. (1991). Local independence and latent class analysis. *Rasch Measurement Transactions*, *5*(3),160.

Borgatti, S. P., Everett, M. G., Freeman, L. C. (2002). Ucinet 6 for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Langeheine, R. & Rost, J. (1988). Latent trait and latent class models. New York: Plenum Press.

Linacre, J. M. (1994). (Dichotomous Mean-square) Chi-square fit statistics. *Rasch Measurement Transactions*, *8*(2), 360.

Linacre, J. M. (2012). A Comment on the HT Person Fit Statistic. *Rasch Measurement Transactions*, *26*(1), 1358.

Linacre, J.M.(1997). An All-Purpose Person Fit Statistic? *Rasch Measurement Transactions*, *11*(3), 582-3.

Sijtsma, K. (1986). A coefficient of deviant response patterns. Kwantitative Methoden, 7, 131-145.

Von Davier, M. (2001). WINMIRA: Latent class analysis, dichotomous and polytomous Rasch models [Computer software]. St. Paul, MN: Assessment Systems Corporation.

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, 2016, so please make your submission by August 1, 2016 for full consideration. Please email Editor\at/Rasch.org with your submissions and/or ideas for future content.

Rasch-related Coming Events

June 16-19, 2016, Thur.-Sat. In-person workshop: Introduction to Rasch measurement analysis in the healthcare sciences and education (in English), Barcelona, Spain (L. Gonzalez de Paz, W. Boone, Winsteps)

July 1-29, 2016, Fri.-Fri. Online workshop: Practical Rasch Measurement – Further Topics (E. Smith, Winsteps), www.statistics.com

July 6-8, 2016, Wed.-Fri. In-person workshop: IRT and CAT using Concerto, Cambridge, UK, www.psychometrics.cam.ac.uk/

July 30-31, 2016, Sat.-Sun. PROMS 2016 Pre-Conference Workshop, Xi'an, China

Aug. 1-3, 2016, Mon.-Wed. PROMS 2016 Conference, Xi'an, China

Aug. 1-Nov. 25, 2016, Mon.-Fri. Online course: Introduction to Rasch Measurement Theory EDU5638 (D. Andrich, RUMM2030), www.education.uwa.edu.au

Aug. 3-5, 2016, Wed.-Fri. IMEKO 2016 TC1-TC7-TC13 Joint Symposium, Berkeley, CA, www.imeko-tc7-berkeley-2016.org

Aug. 12-Sept. 9, 2016, Fri.-Fri. Online workshop: Many-Facet Rasch Measurement (E. Smith, Winsteps), www.statistics.com

Sept. 2-Oct. 14, 2016, Fri.-Fri. Online workshop: Rasch Applications, Part 1: How to Construct a Rasch Scale (W. P. Fisher, Jr.), www.statistics.com

Sept. 28-30, 2016, Wed.-Fri. In-person workshop: Introductory Rasch (M. Horton, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Oct. 3-5, 2016, Wed.-Fri. Intermediate Rasch (M. Horton, Tennant, RUMM), Leeds, UK

Oct. 6-7, 2016, Thur.-Fri. In-person workshop: Advanced Rasch (M. Horton, RUMM), Leeds, UK

Oct. 14-Nov. 11, 2016, Fri.-Fri. Online workshop: Practical Rasch Measurement – Core Topics (E. Smith, Winsteps), www.statistics.com