Spring 2019, VOL. 32 NO. 2; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

Notes:

Transactions of the Rasch Measurement SIG
American Educational Research Association

Overview of The Issue

In this issue of RMT, we have included a few announcements, as well as several research notes.

First, we have included a list of upcoming presentations related to Rasch measurement theory at the annual meetings of the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME), which will be held in Toronto, Ontario, Canada between April 4 and April 9, 2019. Readers can use this list to plan their time at these upcoming conferences.

Second, continuing from the last issue, we have included a summary of another paper that was presented at International Objective Measurement Workshop (IOMW) held in April 2017, prior to AERA and NCME.

Next, we have included two research notes related to person estimation and validity evidence within the framework of Rasch measurement theory.

The last two notes in this issue are related to the community of Rasch scholars. Chien and Shao have provided a note related to citations of Rasch scholars. To finish up the issue, Richard Smith has provided a history of key events in the history of Rasch measurement, with an invitation to readers to suggest additions.

Our plans for the future of RMT are still in progress, but we expect to contact you soon with a survey to get your feedback and suggestions.

Sincerely,

Your RMT Co-editors, Leigh and Stefanie

Rasch Measurement Transactions

www.rasch.org/rmt

Editors: Leigh M. Harrell-Williams
& Stefanie A. Wind
Email submissions to:
Leigh.Williams@memphis.edu or swind@ua.edu
RMT Editors Emeritus: Richard M. Smith, John M. Linacre, &
Ken Royal
Rasch SIG Chair: Hong Jiao
Secretary: Cari F. Herrmann-Abell
Treasurer: Matt Schulz
Program Chairs: Eli Jones & Trent Haines

List of Upcoming Conference Presentations related to Rasch Measurement Theory at AERA and NCME

AERA

Business Meeting:

- Rasch SIG business meeting
  - *Time & location:*
    - Mon, April 8, 6:35 to 8:05pm
    - Fairmont Royal York Hotel, Convention Level, Salon A

Paper Sessions:

- Applications of Rasch models in Educational Assessments
  - *Time & location:*
    - Sun, April 7, 8:00 to 9:30am,
    - Fairmont Royal York Hotel, Mezzanine Level, Tutor 8
  - *Papers:*
    - Examining an observation instrument for mathematics intervention instruction – *Angela Rae Crawford, Yuzhu Zheng, Evelyn Johnson, & Laura Moylan – Boise State University*
    - Examining measurement invariance with multilevel behavioral Rasch models: Does gender affect international food insecurity measures? – *Jue Wang, University of Miami; Victoria Tanaka, George Engelhard, University of Georgia; & Matthew Rabbitt – The US Department of Agriculture Economic Research Service*
    - Measuring inferential integrative reasoning using modern objective measurement – *Alexander Mario Blum, James M. Mason, Jinho Kim, & David Pearson, University of California – Berkeley*
    - Using the Rasch partial credit model to inform the process of teacher performance evaluation – *Richard G. Lambert & Cary Butts – University of North Carolina at Charlotte*
- Methodological considerations of Rasch models
  - *Time & location:*
    - Mon. April 8, 2:15 to 3:45 pm
    - Fairmont Royal York Hotel, Mezzanine Level, Saskatchewan
  - *Papers:*
    - An examination of sensitivity to measurement error of Rasch residual-based fit statistics – *Noah Padgett & Grant B. Morgan, Baylor University*
    - Differential item functioning analysis of a statewide visual arts assessment using a two-stage procedure – *Kelvin Terrell Pompey, Ning Jiang, Yin Burgess, Ashlee A. Lewis, & Jingtong Dou, University of South Carolina*
    - Exploring the impact of score resolution on person fit and

decision consistency in rater-mediated assessments – *Stefanie A. Wind, University of Alabama, A. Adrienne Walker, Gwinnett County Public Schools, Cheng Hua & Abdullah Asilkalkan, University of Alabama*

- Test administration mode effects of hands-on-based test versus paper-and-pencil-based test – *Shaohui Chi & Xiufeng Liu, University at Buffalo – SUNY, & Zuhao Wang, East China Normal University*

Other Rasch-Related Papers in Sessions:

- Advancements in human scoring: Bringing together multiple approaches
  - *Time & location:*
    - Fri, April 5, 12:00 to 2:00pm
    - Fairmont Royal York Hotel, Mezzanine Level, Confederation 5
  - *Paper:*
    - Investigating human essay rating quality in a large-scale assessment using many-facet Rasch measurement – *Xiuyuan Zhang, The College Board*
- Symposium on recent developments in medical certification score reporting
  - *Time & location:*
    - Tue, April 9, 8:00 to 9:30am
    - Metro Toronto Convention Centre, 200 Level, Room 201F
  - *Paper*:
    - Adapting the zone of proximal development for reporting exam results – *Thomas R. O'Neill & Michael R. Peabody, American Board of Family Medicine*
- Investigating course modality, motivation, engagement, self-efficacy, and self-directed learning
  - *Time & location:*
    - Sat, April 6, 2:15 to 3:45pm
    - Metro Toronto Convention Centre, 600 Level, Room 606
  - *Paper:*
    - Evaluating the self-directed learning scale measuring graduate student online learning perceptions: A Rasch analysis approach – *Jian Su, The University of Tennessee, Knoxville & Hongwei Yang, University of West Florida*
- Assessment in international schools
  - *Time & location:*
    - Tue, April 9, 12:20 to 1:50pm
    - Fairmont Royal York Hotel, Convention Level, Salon B
  - *Paper:*
    - Measurement invariance of the learning and study strategies inventory ii for gender and discipline in Egyptian students – *Mohammed Abdelhady Abdelsamea, South Valley University & Ernest C. Davenport, Jr., University of Minnesota*
- Early childhood language and literacy
  - *Time & location:*

        - Sat, April 6, 12:20 to 1:50pm
        - Metro Toronto Convention Centre, 200 Level, Room 202D
    - *Paper:*
        - Rasch modeling of letter sound production in early childhood – *Luke S. Duesbery, San Diego State University*
- Next generation science standards assessment design across use cases: From formative classroom assessment to large-scale accountability assessment
    - *Time & location:*
        - Tue, April 9, 10:25 to 11:55am
        - Fairmont Royal York Hotel, Mezzanine Level, Confederation 3
    - *Paper:*
        - Scoring and dimensionality in next generation science standards assessment – *Courtney Castle, The Woodrow Wilson Academy of Teaching and Learning*
- Supporting children's early learning and kindergarten goals with a standards-aligned assessment: Effective practices in statewide aggregate reporting
    - *Time & location:*
        - Tue, April 9, 2:15 to 3:45pm
        - Metro Toronto Convention Centre, 200 Level, Room 203B
    - *Paper:*
        - Multidimensional item response theory analysis to support formative, longitudinal assessment of early childhood development – *Joshua Syssman & Perman Gochyyev, University of California, Berkeley*
- Emergent methodological considerations for higher education outcomes
    - *Time & location:*
        - Sun, April 7, 11:50am to 1:20pm
        - Sheraton Centre Toronto Hotel, Second Floor, Kent
    - *Paper:*
        - Integrative learning: Development of a measure – *Ethan W. Youngerman, New York University, Laura Stilz Dhal & Mattthew J. Mayhew, The Ohio State University*
- Forming lives of meaning and purpose: A 21st century mission for schools of education?
    - *Time & location:*
        - Sun, April 7, 8:00 to 9:30am
        - Metro Toronto Convention Centre, 700 Level, Room 717A
    - *Paper:*
        - The BC-LAMP portfolio project – Larry Ludlow et al., Boston College
- Science teaching and learning SIG paper session: Problem solving and inquiry
    - *Time & location:*

        - Tue, April 9, 2:15 to 3:45pm
        - Metro Toronto Convention Centre, 700 Level, Room 711
    - *Paper:*
        - A chameleon effect of inquiry-based science teaching on science achievement: Evidence from PISA 2015 – Xian Wu, University of Kentucky
- Selection and Prediction
    - *Time & location:*
        - Sat, April 6, 2:15 to 3:45pm
        - Metro Toronto Convention Centre, 200 Level, Room 201F
    - *Paper:*
        - Establishing a quality control for residency applicant scores – *Jed Wolpaw et al., Johns Hopkins University*

Poster Sessions:

- Rasch Measurement in Education Settings
    - *Time & location:*
        - Tue, April 9, 10:25 to 11:55am
        - Metro Toronto Convention Centre, 300 Level, Hall C
    - *Posters:*
        - Differential item functioning among English language learners on a large-scale mathematics assessment – *Ruixue Liu & Kelly D. Bradley, University of Kentucky*
        - Rasch analysis of puppet interview scales of competence in and enjoyment of science with kindergarteners – *Courtney Donovan, University of Colorado – Denver, Sarah Brenkert, & Maggie Miller*
- Graduate student issues committee graduate student poster session 1
    - *Time & location:*
        - Sat, April 6, 12:20 to 1:50pm, Fairmont Royal York Hotel, Main Level, Imperial Room
    - *Poster:*
        - Effects of probability threshold choice on adjustment for guessing with Rasch modeling – *Tom Waterbury & Christine Demars, James Madison University*
- MTCC Poster Session
    - *Time & location:*
        - Sat, April 6, 8:00 to 9:30am
        - Metro Toronto Convention Centre, 300 Level, Hall C
    - *Poster*:
        - Validating the principal preparation programs policy (4ps) instrument: new evidence from Rasch analysis – *Nahed Abdelrahman, Beverly J. Irby, Lixia Qin, Rafael Lara-Alecio, & Fuhui Tong, Texas A&M University*
- Interactive stitch sampler of equitable learning and teaching with e-textiles in K–12 education
    - *Time & location:*
        - Sat, April 6, 2:15 to 3:45pm
        - Metro Toronto Convention Centre, 800 Level, Room 801A
    - *Poster:*

        - Can we compare teaching behavior across national contexts? Rasch modeling and differential item functioning approach – Ridwan Maulana, University of Gronigen et al.

Roundtable Sessions:

- Topics in Rasch Measurement
  - *Time & location:*
    - Sat, April 6, 10:25 to 11:55am,
    - Fairmont Royal York Hotel, Convention Floor, Concert Hall
  - *Papers:*
    - A Rasch analysis of the teacher readiness scale – *Riza Memis, Ilker Soyturk, Aryn. C. Karpinski, Kent State University*
    - Assessing the measurement properties of school leadership dimension of revised school culture element questionnaire: A Rasch modeling approach – *John K. Rugutt, Illinois State University; Caronline C. Chemosit, Lincoln College; Mohamed A. Nur-Awaleh, Illinois State University*
    - Revising a measure of high school students' mathematics anxiety – *Kelsey Klein, Boston College*
- MTCC poster session
  - *Time & location:*
    - Fri, April 5, 4:20 to 5:50pm
    - Metro Toronto Convention Centre, 300 Level, Hall C
  - *Papers:*
    - Many-facet Rasch analysis of instrument measuring physics teachers' formative assessment knowledge of force and motion – *Marilyn Maxwell Stephens & Dennis Sunal, The University of Alabama*
    - Evidence of fairness in accountability systems: A comparative study of multilevel differential item functioning frameworks – *Elizabeth Adele Patton, University of North Carolina - Greensboro*
- Psychometric methodologies for test validity research and evaluation
  - *Time & location:*
    - Sat, April 6, 12:20 to 1:50pm
    - Fairmont Royal York Hotel, Convention Floor, Concert Hall
  - *Paper:*
    - Investigating the psychometric properties of the classroom assessment literacy inventory for preservice teachers – *Kelli Qua, Case Western University, Aryn C. Karpinski, Kent State University*
- Implementation of protocols in classroom observation
  - *Time & location:*
    - Sat, April 6, 4:10 to 5:50pm
    - Metro Toronto Convention Centre, 700 Level, Room 709
  - *Paper:*
    - Raters' use of an observation protocol for mathematics intervention instruction – *Angela Rae Crawford, Yuzhu Zheng,*

Evelyn Johnson, & Laura Moylan, Boise State University

- Advances in measuring dimensions of teachers and teaching
  - *Time & location:*
    - Fri, April 5, 12:00 to 1:30pm
    - Metro Toronto Convention Centre, 800 Level, Hall G
  - *Paper:*
    - Developing a measurement instrument for teacher key competencies – *Jing Lin, Beijing Normal University, Xiufeng Liu, University at Buffalo – SUNY, Chun-Yeng Chang, National Taiwan Normal University, Tianying Sun, Beijing Normal University*
- Examining and facilitating student self-efficacy: Experiences of marginalized populations
  - *Time & location:*
    - Sat, April 6, 10:25 to 11:55am
    - Sheraton Centre Toronto Hotel, Lower Concourse, Osgoode Ballroom
  - *Paper*:
    - Parenting and school climate predictors of Australian aboriginal students' optimism, coping, self-efficacy, and academic achievement – Helen Joanna Boon, James Cook University

NCME

Paper Sessions:

- Equating: Applications and insights
  - *Time & location:*
    - Sat, April 6, 4:10 to 6:10pm
    - Fairmont Royal York Hotel, Convention Level, Salon B
  - *Papers:*
    - Impact of Rasch item parameter drift in small samples over multiple administrations – *Jason P. Popp & Andrew Jones, American Board of Surgery*
    - Rasch versus classical equating in the context of small sample sizes – *Ben Babcock, The American Registry of Radiologic Technologists & Kari Hodge, NACE International Institute*
- New learning in item analysis research
  - *Time & location:*
    - Sun, April 7, 5:05 to 6:35pm, Fairmont Royal York Hotel
    - Mezzanine Level, Alberta
  - *Papers:*
    - Positive intercultural adaptation: Item weighting and differential item functioning – *Travis Henry, Pedro R. Portes, Ruben Atilano, & Diego Boada Beltran, University of Georgia*
    - Anchors aweigh: How the choice of anchor items affects Rasch vertical scaling – *Tom Waterbury & Christine Demars, James Madison University*

- Fairness issues in test construction
  - *Time & location:*
    - Sat, April 6, 12:20 to 1:50pm
    - Fairmont Royal York Hotel, Mezzanine Level, Manitoba
  - *Paper*:
    - Test construction and selection bias: An investigation using the Rasch model – *Andrew Jones & Jason P. Kopp, American Board of Surgery, Thai Ong, James Madison University*
- Technical considerations in calculating and evaluating reliability
  - *Time & location:*
    - Mon, April 8, 8:00 to 10:00am
    - Fairmont Royal York Hotel, Convention Level, Salon B
  - *Paper:*
    - Examining rating designs with cross-classification multilevel Rasch models – *Jue Wang, University of Miami, Zhenqui Lu, George Engelhard, & Allan S. Cohen, University of Georgia*
- Advances in the evaluation of item response theory models
  - *Time & location:*
    - Sat, April 6, 8:00 to 10:00am
    - Fairmont Royal York Hotel, Mezzanine Level, Quebec
  - *Paper:*
    - Exploring psychometric models for process data from computer-based simulations – *Yanyan Tan, University of Georgia, Matthias Von Davier & Polina Harik, National Board of Medical Examiners*
- Examining impacts of rater effects
  - *Time & location:*
    - Mon, April 8, 8:00 to 10:00am
    - Fairmont Royal York Hotel, Mezzanine Level, Territories
  - *Paper:*
    - Combined effects of rater misfit and differential rater functioning in performance assessments – *Wenjing Guo & Stefanie A. Wind, University of Alabama*
- Advances in evaluating psychometric models
    - *Time & location:*
    - Mon, April 8, 8:00 to 10:00am,
    - Fairmont Royal York Hotel, Mezzanine Level, Manitoba
  - *Paper:*
    - Anchoring rater effects from a suboptimal judging plan: A sensitivity analysis – *Christopher T. Moore, Minneapolis Public Schools*
- Challenges, issues, and opportunities in interrater reliability
  - *Time & location:*
    - Mon, April 8, 4:10 to 6:10pm
    - Fairmont Royal York Hotel, Mezzanine Level, Alberta
  - *Paper:*
    - Rater consistency with a teacher observation protocol – Evelyn Johnson, Yuzhu Zhang, Angela Rae Crawford, & Laura Moylan, Boise State University

- Important test administration and scoring considerations
  - *Time & location:*
    - Mon, April 8, 4:10 to 6:10pm
    - Fairmont Royal York Hotel, Mezzanine Level, British Columbia
  - *Paper:*
    - The effects of test familiarity on person-fit and aberrant behavior – *Hotaka Maeda, University of Wisconsin-Milwaukee & Xiaolin Wang, NBOME*
- New directions in Item Response Theory
  - *Time & location:*
    - Sun, April 7, 12:10 to 1:40pm
    - Fairmont Royal York Hotel, Mezzanine Level, Territories
  - *Paper:*
    - Using the discontinuation rule to reduce the effect of random guessing – *Tianshu Pan, Pearson & Youngmi Cho, American Institutes for Research*
- New insights in differential item functioning analysis
  - *Time & location:*
    - Mon, April 8, 10:25 to 11:55am
    - Fairmont Royal York Hotel, Convention Level, Salon B
  - *Paper*:
    - Evidence of fairness in multilevel data: A comparative study of three differential item functioning frameworks – *Elizabeth Adele Patton, University of North Carolina – Greensboro*

Electronic Board Sessions:

- Electronic board session 1
  - *Time & location:*
    - Sat, April 6, 10:25 to 11:55am
    - Fairmont Royal York Hotel, Main Level, Imperial Room
  - *Paper:*
    - The identification of latent class membership in the mixture Rasch model – *Tongyun Li, Educational Testing Service, Ming Li, Georgetown University, & George Macready, University of Maryland*

List of Recent Publications in Journal of Applied Measurement

Vol. 19, No. 4, Winter 2018

Hierarchical and Higher-Order Factor Structures in the Rasch Tradition: A Didactic – *Perman Gochyyev and Mark Wilson*

Factor Structure of the Community Reintegration of Service-Members (CRIS) in Veterans with Blast-Related Mild Traumatic Brain Injury – *J. Kay Waid-Ebbs, Pey-Shan Wen, David P. Graham, Kathleen Ray, Audrey J. Leroux, Maureen K. O'Connor, and Drew Helmer*

Examination of Item Quality in a State-Wide Music Assessment Program using Rasch Methodology – *Yin Burgess, Jin Liu, and Mihaela Ene*

Validation Instrument to Evaluate Students' Perception of Virtual Manipulatives in Learning Mathematics – *Fereshteh Zeynivandnezhad*

Psychometric Properties and Convergent Validity of the Chinese Version of the Rosenberg Self-Esteem Scale – *Meng-Ting Lo, Ssu-Kuang Chen, and Ann A. O'Connell*

Rasch Analysis of the Revised Two-Factor Study Process Questionnaire: A Validation Study – *Vernon Mogol, Yan Chen, Marcus Henning, Andy Wearn, Jennifer Weller, Jill Yielder, and Warwick Bagg*

A Measurement Model of City-Based Consumer Patriotism in Developing Countries: The Case of Vietnam – *Ngoc Chu Nguyen Mong and Trong Hoang*

Vol. 20, No. 1, Spring 2019

The Effects of Probability Threshold Choice on an Adjustment for Guessing using the Rasch Model – *Glenn Thomas Waterbury and Christine E. DeMars*

Quantifying Item Invariance for the Selection of the Least Biased Assessment – *W. Holmes Finch, Brian F. French, and Maria E. Hernandez Finch*

Rasch Model Calibrations with SAS PROC IRT and WINSTEPS – *Ki Cole*

Student Perceptions of Grammar Instruction in Iranian Secondary Education: Evaluation of an Instrument using Rasch Measurement Theory – *Stefanie A. Wind, Behzad Mansouri, and Parvaneh Yaghoubi Jami*

Computer Adaptive Test Stopping Rules Applied to the Flexilevel Shoulder Functioning Test – *Trenton J. Combs, Kyle W. English, Barbara G. Dodd, and Hyeon-Ah Kang*

Examining Rater Judgements in Music Performance Assessment using Many-Facets Rasch Rating Scale Measurement Model – *Pey Shin Ooi and George Engelhard, Jr.*

Examining Differential Item Functioning in the Household Food Insecurity Scale: Does Participation in SNAP Affect Measurement Invariance? – *Victoria T. Tanaka, George Engelhard, Jr., and Matthew P. Rabbitt*

Accuracy and Utility of the AUDIT-C with Adolescent Girls and Young Women (AGYW) Who Engage in HIV Risk Behaviors in South Africa – *Tracy Kline, Corina Owens, Courtney Peasant Bonner, Tara Carney, Felicia A. Browne, and Wendee M. Wechsberg*

Richard M. Smith, Editor
Journal of Applied Measurement
www.jampress.org

Featured IOMW Plenary Session Paper Summary:

Overview of "Using Guttman errors to explore rater fit in rater-mediated performance assessments"

I presented this manuscript as part of the opening session at IOMW in Spring 2018. In the presentation and paper, I discussed how *Guttman errors* provide useful information for evaluating ratings in rater-mediated performance assessments. For dichotomous items, Guttman errors occur when a test-taker provides a correct response to a difficult item in combination with an incorrect response to a difficult item. Guttman errors for raters are a bit more complicated to describe, but they are based on the same principle. To illustrate, consider a pair of raters *i* and *j* who are ordered such that Rater *i* is more severe (gives lower ratings more often) and Rater *j* is more lenient (gives higher ratings more often)—that is Rater *i* < Rater *j*. A Guttman error would occur if Rater *i* (more severe) gave a higher rating than Rater *j* (more lenient). I provided a full illustration of this point in the manuscript.

In this study, I used simulated and real data to examine the relationship between summaries of Guttman errors (scalability coefficients) and Rasch indicators of model-data fit (Outfit Mean

Square Error statistics). I observed that the two approaches provide similar but not exactly the same information. My major conclusion was that it is possible to use simple nonparametric fit statistics to explore many of the same issues as Rasch fit statistics in rater-mediated assessments.

I have provided the abstract for my manuscript below. The final version of the paper is available in open-access format from *Methodological Innovations:* https://doi.org/10.1177%2F2059799118814396

Abstract

Model-data fit indices for raters provide insight into the degree to which raters demonstrate psychometric properties defined as useful within a measurement framework. Fit statistics for raters are particularly relevant within frameworks based on invariant measurement, such as Rasch measurement theory and Mokken scale analysis. A simple approach to examining invariance is to examine assessment data for evidence of *Guttman errors*. I used real and simulated data to illustrate and explore a nonparametric procedure for evaluating rater errors based on Guttman errors, and to examine the alignment between Guttman errors and other indices of rater fit. The results suggested that researchers and practitioners can use summaries of Guttman errors to identify raters who exhibit misfit. Further, results from the comparisons between summaries of Guttman errors and parametric fit statistics suggested that both approaches detect similar problematic measurement characteristics. Specifically, raters who exhibit many Guttman errors tended to have higher-than-expected Outfit *MSE* statistics and lower-than-expected estimated slope statistics. I discuss implications of these results as they relate to research and practice for rater-mediated assessments.

Stefanie A. Wind
The University of Alabama

Finding Person's Thresholds in Rating Scale Analysis using the Rasch Model

The well-known Andrich or Masters models are the most suitable tools to make measures using rating scale items, in comparison to models of ordered choices (Greene & Hensher, 2010; Hensher, Rose & Greene, 2005). The former are based on the Rasch model and the concept of thresholds between categories (Andrich, 1978, 1978, 1998; Wright & Masters, 1982) while the latter use other type o approaches like Probit or bayesian models (Linacre, 2003).

Andrich (1998) says: "The threshold estimates are independent of the person distribution - so we can tell if there is something wrong with our instrument independently of the distribution.", while Greene & Hensher objection (2010) say: “the thresholds are specific to the person… one of the admitted unrealistic assumptions in many applications is that these threshold values are the same for all individuals.” From Andrich and the Rasch model point of view, once the item and the test fits the Rasch model, then all persons respond to the same construct and the categories are ordered in the same way for everybody; in fact the model provides ordered and clear thresholds between categories. As a consequence, if the persons do not understand the rating scale or have a different way to see the trait to be scored, then their responses will not fit the Rasch model and their answers could be discarded from the set, otherwise the researcher has to make some other provisions to use the data. The definition of thresholds in those approaches is out of the purpose of this paper, but I want to answer two questions:

(1) Can we prove the Greene & Hensher objection (GHO) in a practical problem?

(2) If the GHO is correct, then is it possible to use the Rasch model to obtain the categories’ thresholds for each person?

To explore the problem, I use a 16-item questionnaire that explores the expectations of graduate students in some topics such as: research areas, professional competencies, opportunities to find a job, and so forth. In this questionnaire, 10 items are 3 categories rating scale (categories codified as 1, 2, 3) and 6 items are dichotomous (codified as 1, 3). The questionnaire was administered to 90 respondents. The questionnaire is analyzed three times as follows.

Analysis 1. In order to analyze the data, the set of answers contains 90 rows for the persons and 16 columns for the items and the control file was used in Winsteps® (Linacre, 2018). The software outcomes include the Wright map (Figure 1), fit statistics and other informative elements.

The scale in the Wright map runs from -2 to 4 logits, the difficulties of the items are centered in 0 logits, and the mean of the persons is in 1.03 logits.

The graph of the category probability curves (Figure 2) shows two distinct Andrich thresholds, for the whole set of items.

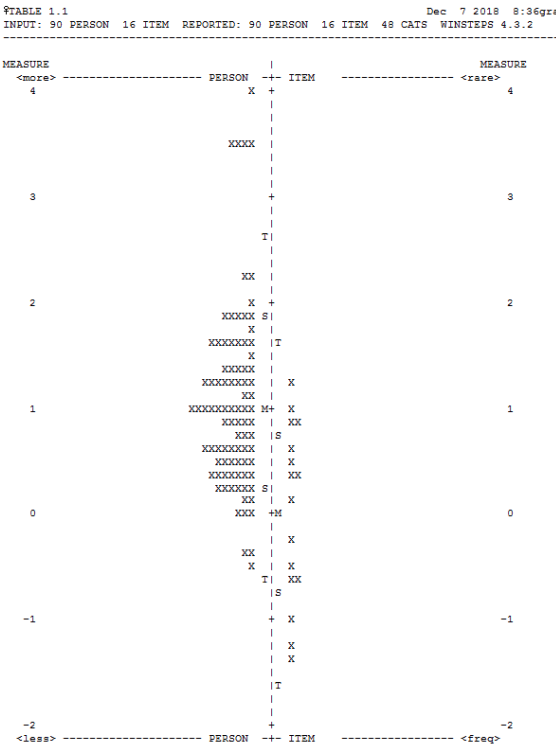

Figure 1. *Wright map for Analysis 1.*

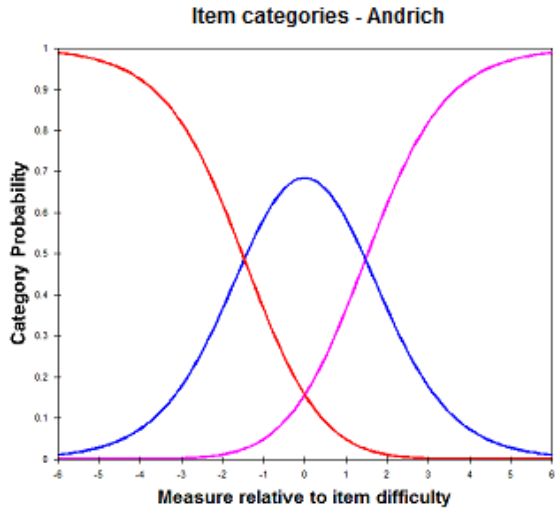

Figure 2. *Category probability curves for Analysis 1.*

Analysis 2. In order to determine the categories' thresholds for the persons, I transpose the response set. Now in the 16 rows there are items and in the 90 columns there are persons and I run a Rasch analysis. Due to the data organization the software will treat the items as persons and *vice versa*.

There are several findings in this approach. First of all, the rating scale of Analysis 2 shows the same scale than Analysis 1, corresponding the mean values of item difficulties and persons measures, but it is evident that the distribution of items and persons in the Wright map (Figure 3) has a different aspect.

In this second analysis, the category probability curves (Figure 4) do not clearly distinguish between the two thresholds, as it was in Analysis 1.

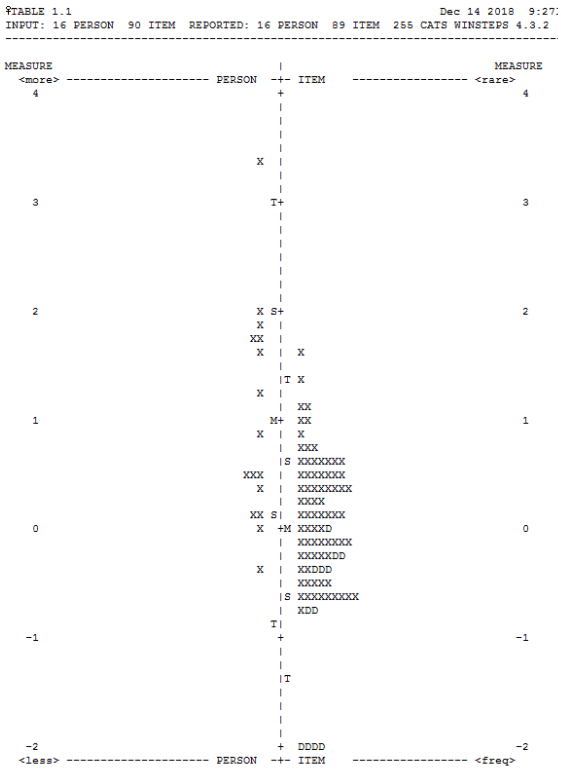

Figure 3. Wright map for Analysis 2.

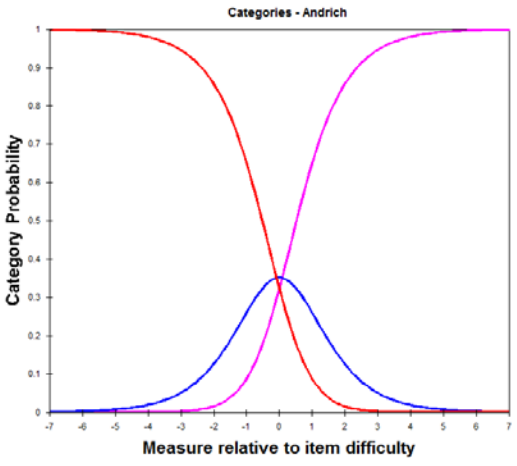

Figure 4. *Category probability curves for Analysis 2.*

Analysis 3. In this final analysis the Winsteps® file was modified including the control ISGROUPS = 0 to use the Masters' partial credit model, where the items (now persons) do not share the same rating scale. Now it is possible to produce the category probability curves of each person (Figure 5a to 5d). It can be seen that persons do not share the same thresholds (as Greene and Hensher suggest!). In fact, every person really has a different approach when selecting between categories 1, 2 and 3.

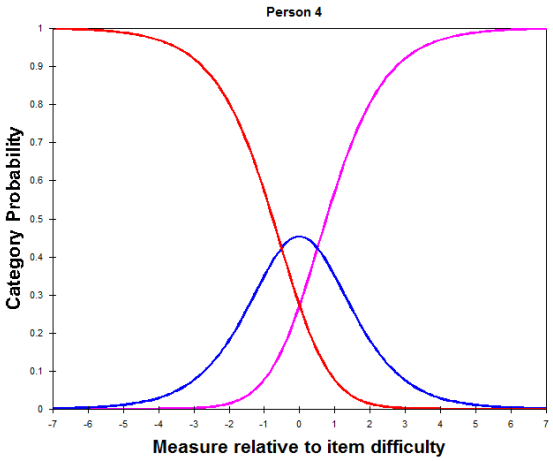

Figure 5a. *Category probability curves for Analysis 3.*

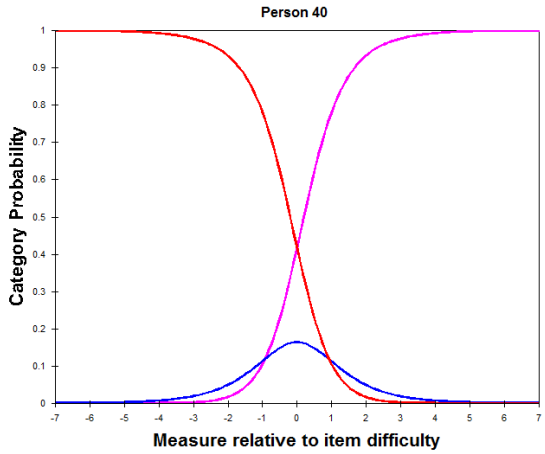

Figure 5b. *Category probability curves for Analysis 3.*

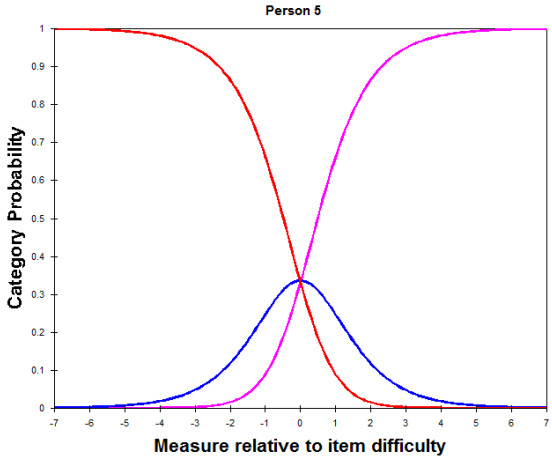

Figure 5c. *Category probability curves for Analysis 3.*

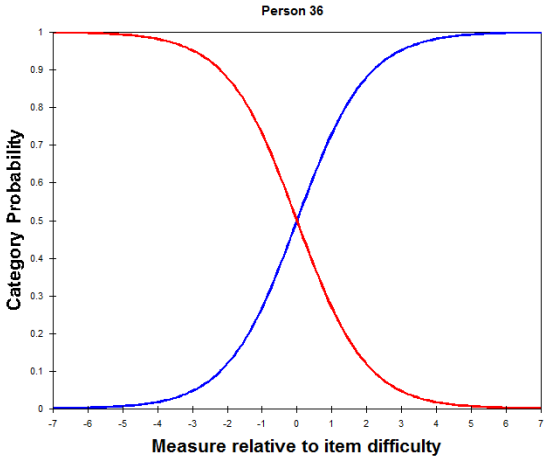

Figure 5d. *Category probability curves for Analysis 3.*

Four patterns can be found in Analysis 3: (1) Two distinct thresholds of the three categories (Person 4 in Figure 5a); (2) two unordered thresholds (Person 40 in Figure 5b); (3) one threshold in a single value for the three categories (Person 5 in Figure 5c) and (4) only one threshold for only two categories (Person 36 in Figure 5d). These four patterns correlate with the expertise and some other professional characteristics of the respondents, but the results are not presented here. A discussion about the thresholds and their implications can be found in Tennant (2004) or Linacre (2004).

There are certainly many other things to say about the use of the transposed set of responses, in particular about its applications on fitting, item-test correlation and factor analysis. But the main original questions were solved: The thresholds are specific to the respondent and it is possible to determine them using the Rasch model.

Agustin Tristan
Instituto de Evaluación e Ingeniería Avanzada. San Luis Potosí, Mexico.

References

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

Andrich, D. (1979). A model for contingency tables having an ordered response classification. Biometrics, 35, 403-415.

Andrich, D. (1995). Distinctive and incompatible properties of two common classes of IRT modules for graded responses. Applied Psychological Measurement. 19, (1) 101-119.

Andrich, D.A. (1998). Thresholds, Steps and Rating Scale Conceptualization.

Rasch Measurement Transactions, 12:3 p. 648-9.
Greene, W.H. & Hensher, D.A. (2010). Modeling ordered choices. A primer. Cambridge University Press. Cambridge, UK.

Hensher, D.A., Rose, J.M. & Greene, W.H. (2005). Applied choice analysis. Cambridge University Press. Cambridge, UK.

Linacre J.M. (2004). Predicting Measures from Rating Scale or Partial Credit Categories for Samples and Individuals, Rasch Measurement Transactions, 18:1 p.972
Linacre, J. M. (2018). Winsteps® Rasch measurement computer program User's Guide. Winsteps.com.
Linacre, J.M. (2003). What is Item Response Theory, IRT? A tentative taxonomy. Rasch Measurement Transactions, 17:2 p. 926-927
Tennant A. (2004). Disordered Thresholds: An example from the Functional Independence Measure, Rasch Measurement Transactions, 17:4 p.945-948
Wright, B.D. & Masters, G.N. (1982). Rating Scale Analysis, Chicago: MESA Press, www.rasch.org/books.htm

A Note on the Rasch Model and the Instrument-Based Account of Validity

Introduction

The instrument-based approach to validity (Borsboom, Mellenbergh, & van Heerden, 2004) states that a test is valid if (1) the construct exists and (2) there is causal relationship between the construct and the test scores. Borsboom and Markus (2013) use the mechanical metaphor to explain validity. They state that a test is like a mechanical instrument such as a thermometer or a bathroom scale. Variations of the construct must *cause* variations in the test scores. In a thermometer, changes in the temperature cause the movements of the mercury level and thus the readings (scores).

This is in contrast to the currently dominant account of validity which argues that validity is the appropriateness of the decisions and interpretations based on the test scores. According to the latter, we do not validate tests but the uses, interpretations, and the decisions we make based on test scores (Messick, 1989). In the argument-based approach to validation (Kane, 2013), in the first step we specify the inferences and uses we want to make based on the test

scores and in the second step we provide evidence or backing to support the proposed inferences and uses. The process of validation is the process of showing that certain interpretations and uses of the test scores are justified.

How does the Rasch model help?

The formulation of the Rasch model (RM) provides a mechanism to check the two conditions for validity as delineated by Borsboom, et al. (2004). When the RM fits, it means that the most important assumption of the model, i.e., conditional independence holds. Conditional independence stipulates that conditioning on the latent trait the item residuals should be uncorrelated. In other words, after the shared variances among the items (observed variables) are captured the unique variances (residuals or errors) should be independent. The latent trait is incorporated in the covariation among the items and when it is extracted the relationship between the items is eliminated. Therefore, it is logical to assume that the latent trait is the cause of their covariations although other reasons cannot be ruled out (Baghaei & Tabatabaee-Yazdi, 2016). This is interpreted as the existence of a construct underlying all the item responses which causes their variations. Therefore, when the RM holds the first condition for validity is satisfied.

As for the second condition, the basic formulation of the Rasch model allows us to check it directly. The item response function for the RM is:

$$P\,(X_{ni}=1) = \frac{e^{\theta n-\beta_i}}{1+e^{\theta n-\beta_i}} \qquad (1)$$

The above function expresses that the probability that person *n* gives a correct reply to item *i* is governed by the person's ability θ_n and item's difficulty β_i. In Equation 1, if θ_n, i.e., the latent trait, increases the probability of a correct response increases and if it decreases the probability of a correct response decreases. In other words, levels of the construct, indicated by the person parameter θ, causally determine the observed score X_{ni}. Therefore, the causal relationship between the latent trait and the test scores can be tested. Note that in the latent trait models the latent variable is linked to the probability of the observed data not the data themselves.

If a test is valid, Equation 1 should correctly predict the probability of a correct response to each item. This can be checked by examining the item characteristic curves and the item fit values (infit and outfit). The item fit values are computed by standardizing and averaging the differences between the model predictions and the actual observed responses. That is why the misfitting items are deleted from the

tests; responses to these items (more accurately the probability of correct response) do not conform to the levels of the latent ability.
One could argue that the classical test theory (CTT) also provides the same mechanism. The item-total correlations are indications of the relationship between the observed item scores and the construct score, if we assume the total score as an indication of the location on the construct. A high positive item-total correlation means that those with a higher location on the variable continuum have higher chances of giving a correct response to the item and vice versa. Items with low or negative item-total correlation do not conform to this condition and are discarded.

The flaw in this argument is that the CTT does not incorporate a latent ability score. The ability score is the summation of the correct responses. It only represents the content of the test. In latent trait models, the latent ability scores transcend the test content and are assumed to be parameterizations of the ability above and beyond the items included in the test. This interpretation is justified because latent trait models are testable. We assume that there is a real entity which causes the item responses and the position of the individuals on this entity can be estimated. When a latent trait model fits it means that the item responses were generated according to the model and we can determine how they would behave. The fit of the model has implications about the joint probability distribution of the item responses and one can estimate the latent ability parameter based on the observed responses, assuming that the responses were generated by the model. "Given that Subject A has value X on the latent variable, A has Probability Distribution Y over the item responses" (Borsboom, Mellenbergh, & van Heerden, 2003, p. 205). Note that the fit of the latent trait model renders the sumscores interpretable as construct scores too.

Even if we adhere to the argument-based account of validity the Rasch model can play a crucial role in validation. Under the argument-based approach one needs to first specify the inferences that one wants to draw based on the test scores (interpretation/use argument) and in the second step provide backing for those inferences (validation argument). A basic inference in test validation is the *inference from test scores to an underlying trait* (Kane, 2013). This inference is almost always assumed when we develop and validate a test, even if not directly broached. As highlighted above, the RM provides the right mechanism and apparatus to support this inference.

Purya Baghaei
Roya Shoahosseini
English Department, Mashhad Branch,
Islamic Azad University, Mashhad, Iran

References

Baghaei, P., & Tabatabaee-Yazdi, M. (2016). The logic of latent variable analysis as validity evidence in psychological measurement. *The Open Psychology Journal, 9*, 168–175.

Borsboom, D., & Markus, K. A. (2013). Truth and evidence in validity theory. *Journal of Educational Measurement, 50*, 110–114.

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

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

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: American Council on Education and Macmillan.

The most cited Rasch Scholars on Pubmed in 2018

Each year in June, millions of scholars pay close attentions to the Journal Citation Reports (JCR) locating journal impact factors (JIF). However, no such personal author impact factors (AIF; Pan & Fortunato, 2014) were applied to individual scholars, even the h-index (Hirsch,2005), g-index (Egghe. 2006) and x-index (Fenner et al., 2018) have been used for measuring author-level metrics baed on both publications and citations of a scientist or scholar.

One of the shortcomings is the assumption of equal credits for all coauthors in an article (Petersen et al., 2014; Sekercioglu, 2008). Many AIF concepts have already proposed before (Chien, 2006; Petersen et al., 2010), but we are not aware of any empirical study that can successfully demonstrate illustrations for quantifying coauthor contributions in the scientific disciplines.

We developed an author-weighted scheme (AWS) based on the Rasch rating scale model (RSM; Andrich. 1978) as Eq.(1) that can be used for computing individual research achievement(IRA) in a discipline.

$$W_{ij} = \frac{\exp(\gamma_{ij})}{\sum_{j=0}^{m}\exp(\gamma_{ij})} = \frac{2.72\text{^}\gamma_{ij}}{\sum_{j=0}^{m}2.72\text{^}\gamma_{ij}}, \quad (1)$$

The sum of authorships equals 1 for each paper. The weights for each author are shown in Table 1 on the first scenario when the powers (γ_{ij}, theta) for the ordered author name (i) and the article (j) are assigned for m to 1, where the author number is m+1. More importance is given to the first (=exp(m), primary) and the last (=exp(m-1), corresponding or supervisory) authors, while it is assumed that the others (the middle authors) have made smaller contributions.

Alternatively, threshold difficulty (delta) approach based on graded response model (Samejima, 1969) as Eq. (2) and (3) can be applied to yield author weights when delta was set from 1 to m in an integer order.

$$W_{ij} = p_{ij} - p_{ij+1}, \qquad (2)$$

$$P_{ij} = \frac{\exp[Da_j(\theta - \delta_i)]}{1+\exp[Da_j(\theta - \delta_i)]} \approx \frac{\exp[(0 - \delta_i)]}{1+\exp[(0 - \delta_i)]} \qquad (3)$$

For instance, if threshold difficulties are from 1 to 2 when three authors exist, the respective probabilities are 0.27 and 0.12 according to Eq(3). The weights for each author are 0.73, 0.15 and 0.12, according to Eq(2), see Table 1 on the middle scenario.

$$P_{ij} \approx \frac{\exp[(0 - (\delta_i - Adj)]}{1+\exp[(0 - (\delta_i - Adj)]}, \qquad (4)$$

Table 1. *Coauthor-weighted credits across types of the designated model*

Model type:		**Rasch model**			**GR model**			**GR-adj model**		
Author number	**1**	**2**	**3**	**4**	**2**	**3**	**4**	**2**	**3**	**4**
Threshold	0	1	2	3	1	2	3	1	2	3
First author	1	0.73	0.67	0.64	0.73	0.73	0.73	0.73	0.69	0.68
Corresponding author		0.27	0.24	0.24	0.27	0.15	0.15	0.27	0.17	0.17
Second author			0.09	0.09		0.12	0.07		0.14	0.09
Third author				0.03			0.05			0.06
Difficulty adj								0.2	0.25	0.3

If the adjustment difficulties (e.g., 0.2,0.25, 0.3, etc.) are applied to Eq (4), the weights for each author are 0.69, 0.17 and 0.14, see Table 1 on the right-hand side scenario, which might be almost equivalent to the result from Rasch model in Table 1 on the lest-side scenario if the adjustment parameters are applied.

A total of 2,232 author names and 528 paper IDs related to Rasch model or Rasch analysis in years from 2015 to 2017 were downloaded from Pubmed on

Aug. 10th, 2018. More than 762 articles were cited to 297 articles within 2015 to 2018. Personal AIFs were thus obtained for each scholar through a series of computation, see the link at https://youtu.be/CJJ-uV8fYls.

Table 2 shows J Appl Meas was ranked at the first placement among journals on the topic of the Rasch model or Rasch analysis from 2015 and 2017.

Table 2. *The top 10 journals included in this study*

Journal on Pubmed	Count	%
J Appl Meas	28	5.30
Health Qual Life Outcomes	20	3.79
Educ Psychol Meas	17	3.22
PLoS One	17	3.22
Qual Life Res	17	3.22
Arch Phys Med Rehabil	15	2.84
Disabil Rehabil	14	2.65
J Rehabil Med	11	2.08
J Clin Epidemiol	7	1.33
Eval Health Prof	6	1.14

Personal AIF for each scholar were shown on a visual map in Figure 1.

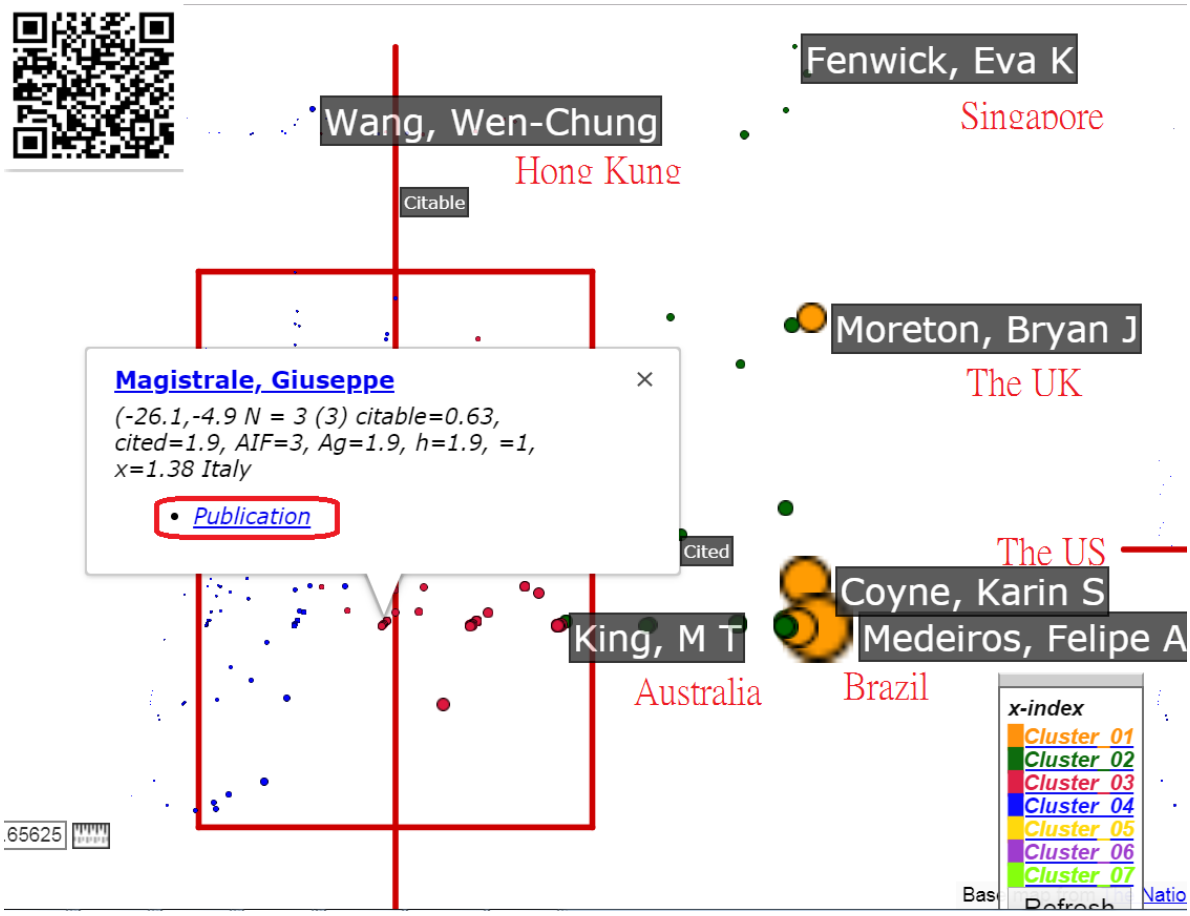

Figure 1. *Personal AIF for each scholar shown on a visualization map*

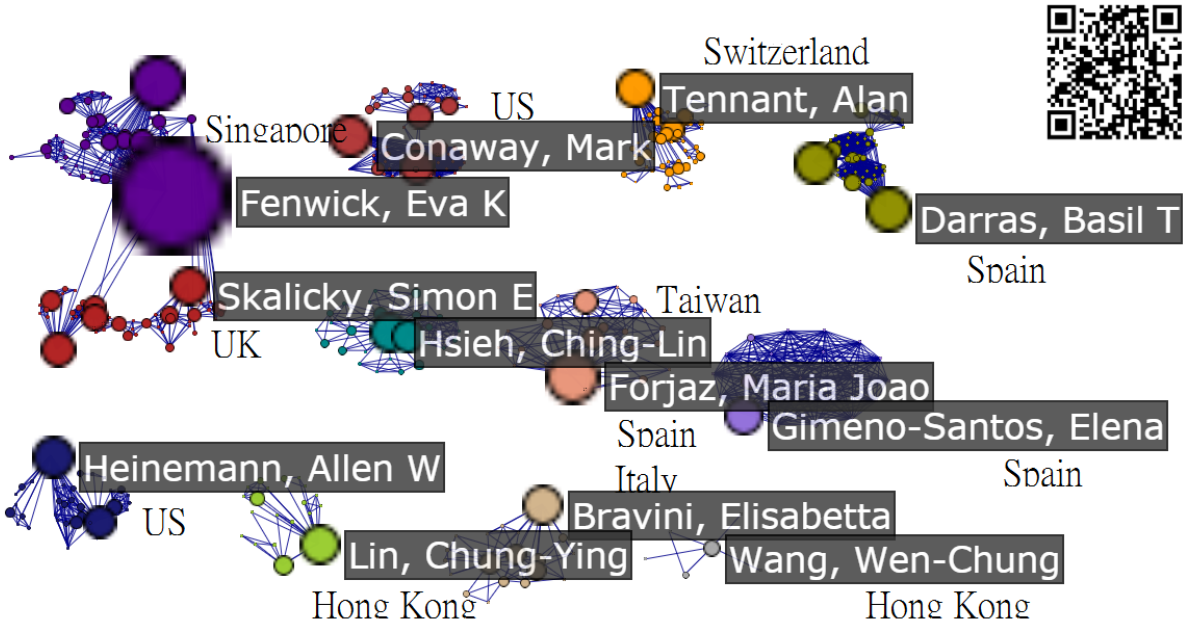

Figure 2. *Author based network related to Rasch analyses on Pubmed*

We can see that the author Felipe A Medeiros from Brazil with an AIF (=26=16.47/0.63, a single citable article with PMID=25444345)(Medeiros, 2015) was colored by a biggest yellow bubble at the right-hand side in Figure 1, indicating he is the most cited Rasch scholar based on Pubmed from 2015 to 2017, followed by Karin S Coyne from the US(19=12.23/0.64) and Elena Gimeno-Santos from Spain (=15=9.48/0.63).

The top three prolific authors on the upmost top in Figure 2 are Eva K Fenwick from Singapore with 5.02 citable articles and x-index = 2.25, Vijaya K Gothwal from India with 2.63 citable articles and x-index = 2.0, and Wen-Chung Wang from Hong Kung with 1.4 citable articles and x-index = 1.15. Interested readers are recommended to scan the QR-code at the left top of Figure 1 to see the dashboard on Google Maps for more details.

The top 12 author clusters separated by social network analysis are shown in Figure 2, on which we see the author Eva K Fenwick from Singapore gain the most highly centered collaborations. Similarly, interested readers are suggested to scan the QR-code at the right top of Figure 2 to see the dashboard on Google Maps for more details.

We used the extended Rasch RSM as an algorithm for computing the contribution weights for each author and performed a sensitivity analysis by the number of coauthors in Table 1 in comparison to the graded response model. Through which, the AIF can be computed and applied to compare the IRA among scientists. As such, the AIF computed by the Rasch baed AWS plays an important

role for scholars, like Thomson Reuters annually releases JIF for the indexed journals, in the discernible future.

Tsair-Wei Chien[1], Yang Shao [2]
Corresponding author: Tsair-Wei Chien
Email: smile@mail.chimei.org.tw

Affiliated institutes:
[1]Chi Mei Medical Center, Taiwan,
[2] Tongji Zhejiang College, Jiaxing, China

References

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

Chien, T. W., Chow, J.C., Chang, Y., Chou, W. (2018). Applying Gini coefficient to evaluate the author research domains associated with the ordering of author names: A bibliometric study. *Medicine, 97*(39), e12418.

Egghe, L.(2006). Theory and practice of the g-index. *Scientometrics, 69*,131-152.

Fenner, T., Harris, M., Levene, M., Bar-Ilan, J.(2018). A novel bibliometric index with a simple geometric interpretation. *PLoS One,13*(7), e0200098

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proc Natl Acad Sci* USA,102,16569-72.

Medeiros, F. A., Gracitelli, C. P., Boer, E. R., Weinreb, R. N., Zangwill, L. M., Rosen, P. N.(2015). Longitudinal changes in quality of life and rates of progressive visual field loss in glaucoma patients. *Ophthalmology,122*(2), 293-301.

Pan, R.K. & Fortunato, S.(2014). Author impact factor: Tracking the dynamics of individual scientific impact. *Sci Rep., 4*, 4880.

Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., Rungi, A., Riccaboni, M., Stanley, H.E., Pammolli, F.(2014). Reputation and impact in academic careers. *Proc Natl Acad Sci U S A, 111*(43),15316-21.

Petersen, A. M., Wang, F., Stanley, H. E. (2010). Methods for measuring the citations and productivity of scientists across time and discipline. *Phys Rev E Stat Nonlin Soft Matter Phys,81*(3 Pt 2), 036114.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*(4, Pt. 2), 100.

Sekercioglu, C. H.(2008). Quantifying coauthor contributions. *Science, 322* (5900), 371.

The Ties That Bind

In the following pages, we have shared Richard Smith's account of key events in the history of Rasch measurement theory.

With this contribution, Richard has invited RMT readers to propose additions to the timeline. Readers who are interested in contributing to the timeline should contact Richard at editor@jampress.org.

Richard M. Smith
Editor, *Journal of Applied Measurement*

Note from RMT co-editors: Leigh & Stefanie added the change of RMT editorship to 2018 on the timeline

The Ties that Bind

	1901	Georg Rasch's birth
	1926	Benjamin Wright's birth
Thurstone & Chave *Measurement of Attitude*	1929	
	1931	L. L. Thurstone appointed Chief Examiner, U of C
Rasch and Ragnard Frisch, Oslo	1934	
Rasch and R. A. Fisher, London		
	1938	Ralph Tyler appointed Chief Examiner, U of C
	1946	Rasch analyzes IQ tests
Rasch and L.J. Savage, Chicago	1947	Cowles Commission
	1948	Wright and Thurstone, U of C
	1949	Wright and Tyler, U of C
	1950	Wright and Bettelheim, U of C
	1952	Rasch analyzes reading tests
	1954	Benjamin Bloom appointed Chief Examiner, U of C
Rasch and Lee Cronbach	1955	Cronbach interviews Rasch for ONR in Copenhagen
	1957	Wright (57) joins faculty at U of C
	1959	Rasch and Frisch
Rasch - *Probabilistic Models* published	1960	Rasch and Wright, U of C (March/June)
Rasch - Berkeley Symposium (June)		
	1961	MESA Special Field created at U of C
	1963	Bruce Chopin (67) arrives at U of C
Bloomers ISU start of three Rasch dissertations	1964	First computer calibration programs
		Wright's first visit to Denmark
		Wright's first Rasch classes at U of C
Jane Loevinger MPA discussant	1965	Midwest Psych. Assoc. mtg.
		Nargis Panchapakesan (69) arrives at U of C
		UCON algorithm developed
		Wright and Chopin visit Rasch
	1966	Bloom and Rasch
Ledyard Tucker, Lou Bashaw	1967	Psychometric Society meeting
Gerhard Fischer and Rasch		ETS Invitational Conference (Bloom organizes program)
Wright - *Sample-free Test Calibration ….*		Choppin rating scale program
Choppin's item bank paper in *Nature*	1968	Rasch teaches at U of C
Dick Woodcock - *Key Math* scoring sheet	1969	First AERA Pre-session
Wright and Panchapakesan *E&PM* article		Rasch gives last lecture at pre-session
Ron Hambleton - Rasch dissertation U of. Toronto		(Attendees include: Angoff, Bashaw, Hambleton, Lenke, Rentz, and Woodcock)
		First MESA Ph.D at U of C
	1971	Ron Mead (76) arrives at U of C
		David Andrich (73) arrives at U of C
		Penny Edgert (75) arrives at U of C
		Joseph Ryan (77, Bloom) arrives at U of C
	1972	Rasch's retirement lecture
		Wright visits NFER and Chopin
		Graham Douglas (74) arrives at U of C
		Christmas flight to London
		CALFIT Program (Edgert)
	1973	Gerhard Fischer, LLTM
		Rasch visits Chicago and Georgia (Rentz)
		AERA Training Session N.O. (CALFIT used)
Otis Dudley Duncan - loglinear publications	1974	Rasch visits UWA in Perth for 6 months
		Portland Public Schools item banks
Rentz - *National Reference Scale for Reading*	1975	

Jim Lunsden - Test Theory in the *Annual Review of Psychology*
Wright - *JEM* Special Issue
Woodcock-Johnson psycho-educational battery
Earling Andersen- *Psychometrika* - sufficiency proof for polytomous model
Ross Lambert VA Blind Center
Andrich - *Psychometrika* - current form of RS and PC models based on Andersen's coefficients
Wright and Stone - *Best Test Design* published
Andrich - *Biometrics* - interprets thresholds
Rasch - *Probabilistic Models* reprinted

Louis Guttman in Chicago
Wright and Masters - *Rating Scale Analysis*
Jim Lumsden
Otis Dudley Duncan Rasch publications
Wright and Bell *JEM* item banking article
Grosse and Wright standard setting
Jack Stenner – development of Lexiles begins
Schulz, Lambert, Becker functional assessment publication

Ross Lambert, Richard Harvey - PECS

Rasch Measurement Transactions Vol. 1
Andrich - *Rasch Models for Measurement*

Wright and Linacre - *APM&R* article

Rasch analysis of *PES*
Carl Granger begins Rasch analysis of *FIM*

OM:TIP Vol.1 published
Rasch analysis of *FIM* published

OM:TIP Vol.2 published

OM:TIP Vol.3 published
Wright & Stone - *Measurement Essentials*
McNamara - *Measuring Second Language Performance*
JOM first issue published
Lexiles commercially available (MetaMetrics)
OM:TIP Vol.4 published

1976 First BICAL Program (Mead)

1977 AERA Training Session N.Y.
Andrich spends 6 months in Denmark

1978 Toronto AERA Presession (Wright, Choppin, Andrich, Mead, Ryan, and Draba)
Robert Draba (78) item bias

1979 Partial Credit model
Andrich interviews Rasch in Denmark
1980 Georg Rasch dies
Geoff Masters (80) rating scales
1981 First IOMW at Univ. of Chicago
1982 Richard Smith (82) fit statistics
IOMW 2 in Perth
1983 Larry Ludlow (83) residual analysis
1984 Mark Wilson (84) hierarchical development
MicroScale program with missing data feature
1985 IOMW3 in Chicago
George Engelhard (85)
Jennifer Bosma (85) CAT in elementary schools
1986 IOMW4 in Chicago
1987 Matthew Schulz (87) functional assessment in rehabilitation
MSCALE program developed
FACETS program developed
Rasch Measurement SIG formed
1988 First SIG Sessions at AERA
William Fisher (88) Truth, method, and measurement
1989 IOMW 5 at Berkeley
Mike Linacre (89) FACETS model
Ray Adams (89) measurement error
Carol Myford (89) judge agreement
1990 Nikolaus Bezruczko (90) aesthetic judgment
BIGSACLE program developed
1991 IOMW 6 in Chicago
Mike Linacre becomes second *RMT* editor
BIGSTEPS program available
1992
1993 IOMW 7 at Atlanta
Quest program available
1994
1995 IOMW 8 at UC Berkeley
1996 First IOMC in Chicago

1997 IOMW 9 in Chicago
Greg Stone (97) Rasch standard setting
First demonstration of RUMM program
ConQuest program available

	1998	Second IOMC in Chicago Winsteps program available
	1999	First Smith and Smith training session
JAM V1, N1 published	2000	JAM indexed in Pub Med & *Index Medicus* IOMW 10 at LSU Medical Center
OM:TIP Vol. 5 published		
Bond and Fox - *Applying the Rasch Model*	2001	Rasch's 100th Birthday party - Copenhagen
	2002	IOMW 11 at LSU Medical Center
	2003	Ben Wright Festschrift at RIC, Chicago
Introduction to Rasch Measurement published	2004	IOMW 12 at James Cook University, Cairns
Quantiles commercially available (MetaMetrics) Wright and Stone - *Making Measures* published *Rasch Measurement in Health Sciences* published		
	2005	First PROMS meeting - Kuala Lumpur (yearly)
Wilson - *Constructing Measures* published *Applications of Rasch Measurement in Science Education* published	2006	IOMW 13 at UC Berkeley - 25 years of IOMW
Lexile Framework for Writing commercially available (MetaMetrics)		
Rasch Measurement: Advanced and Specialized Applications published	2007	
Humphry - *The implied unit in the Rasch Model* 2PL special case of Rasch model	2008	IOMW 14 at NYU
Criterion Referenced Testing published	2009	Improving Efficiency in Health Outcome Outcomes Conference - Chicago 40 years of Rasch training sessions
Advances in Rasch Measurement, Vol. 1 published	2010	50th Anniversary of Rasch's book - Copenhagen IOMW 15 at Univ. of Colorado
Advances in Rasch Measurement, Vol. 2 published	2011	17th Smith and Smith training session
	2012	IOMW 16 in Vancouver, Canada 25 years of Rasch SIG sessions 25 years of RMT Eighth PROMS meeting – Jia Xing Kenneth Royal becomes third RMT editor
Engelhard - *Invariant Measurement* published	2013	
	2014	IOMW 17 in Philadelphia
	2015	IOMC2015 in Chicago (Coffee with Ben, April) Ben Wright dies (October)
	2016	IOMW 18 in Washington, DC
	2017	IOMC 2017 in Chicago at SRAL
	2018	IOMW 19 in New York Stefanie Wind & Leigh Harrell-Williams become RMT co-editors (4th editor)