FALL 2020, VOL. 33 NO. 3; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

Transactions of the Rasch Measurement SIG
American Educational Research Association

Overview of The Issue

In this issue of RMT, we have included one research note and several announcements that may be of interest to the Rasch community.

The issue begins with a peer-reviewed research note related to Rasch measurement and graph theory. We thank our peer-review volunteers for their valuable comments on this research note.

In SIG news, we have included a call for nominations for the Senior Scholar Award. We encourage readers to consider submitting nominations for this important award.

Next in the issue is an announcement about a new YouTube channel that may be of interest to our readers, followed by a note about software. Then, we provide an announcement about a new book on Rasch measurement theory.

The issue rounds out announcements about upcoming Rasch courses that may be of interest to our readers.

As always, we welcome your contributions to the next issue for RMT. Please contact us at the email address below if you wish to submit something for inclusion.

Sincerely,

Your RMT Co-editors, Leigh and Stefanie

Rasch Measurement Transactions
www.rasch.org/rmt

Rasch Measurement and Graph Theory

Nikolaus Bezruczko[1]
Jiaqi Mao[2]
Kyle Perkins [3]

[1]Department of Clinical Psychology, Chicago School of Professional Psychology

[2]Department of Psychology, Sun Yat-Sen University

[3]Retired Professor, Florida International University

Graphs are mathematical structures used to model pairwise relations between objects. A graph is an ordered pair consisting of a set of vertices (also called nodes or points) and a set of edges (also called links or lines). Graph theory, a mathematical method, has been used to conceptualize and to quantify the topological structure of complex networks (Bullmore and Sporns, 2009; Diestel, 2017). For example, graph theory has been utilized in engineering to design circuit connections (Banasode and Umathar, 2018), in computer science to study algorithms (Riaz and Ali, 2011; Shirinivas, Vetrivel, and Elango, 2010), in linguistics to study the syntax of the hierarchical branching tree parsing of head-driven phrase structure grammar (Piperski, 2015), and in neuroimaging to examine the functional abnormalities in individuals with reading disability (Horowitz-Kraus, Buck, and Dorrman, 2016; Edwards, Burke, Booth, and McNorgan, 2018; Yang and Tan, 2019; Mao, Liu, Perkins, and Cao, under review).

While less expansive in social research, Tatsuoka (1986) reported graph theory applications to order-theoretic studies of dimensionality, as well as hierarchical structures, and cluster analysis. In addition, Garner and Engelhard (2000) examined corollaries between graph theory and Rasch models, while Wright and Stone (1979) presented item bank networks based on graph theory principles. Garner (2002) also emphasized the usefulness of graph theory for analyzing assessment networks. In general, these applications emphasized demonstrating links among test items and test forms to gain insights into continuity of scale structures. The purpose of this paper, however, departs substantially from those studies by shifting attention from scale concerns to appraisal of the entity being measured, which addresses Michell's (2004) claim that "mainstream psychometricians have neglected to test the hypothesis that psychological attributes are quantitative" (p.121). Michell (2003, 2005, 2013) has also offered a hypothesis that a quantitative attribute is the basis of measurement and is a class of properties or relations. In addition, in order to be measurable, attributes must be continuous quantities.

Rasch models assume their linear representations are qualitatively meaningful. Representing a critical collective view, Michell rejects that claim, while dismissing relevance of an explicit probabilistic concatenation procedure implemented during scale construction to satisfy empirical requirement for demonstrating measurability. This critical view contends that ability variation in nature is qualitatively discrete, an explicit example of a nonphysical entity and its quantitative ordering is discrete; hence, Rasch model allusions to mathematical continuity are formally rejected, as well as claims that Rasch models conform to fundamental measurement. According to this critical view, Rasch models provide a mathematical representation of idealized scientific measurement, but they do not address ontological issues. This discussion, of course, is long standing and unresolved, while a voluminous and contentious literature continues to surround it (but see Reese, 1943). Moreover, this line of criticism was renewed in a recent attack. Contemporary "psychometrics is a fashionable fraud". "Hence, presenting psychological tests as instruments of measurement is a fraud" (Michell, p.17, 2020).

In the present report, our intent was to demonstrate quantitative continuity of observations on a Rasch calibrated dimension reported in Bezruczko, Frank, and Perkins (2016) with graph theory. Treatment of items as nodes and connections between items as edges contingent on passing items provides a context for demonstrating continuity independent of the instrument's estimation procedure. Moreover, these connections among observation should define an overall network that maintains coherence even as the measurement model implements scale location parameterization. Demonstrating continuity among discrete quantities outside the measurement process, then, provides concrete justification for the probabilistic simultaneous conjoint concatenation implemented during Rasch model item and person calibration. An assertion in this report is graph theory confirmation of nodal connections among discrete quantities satisfies the measurability hypothesis presented above.

This measurability strategy was addressed with data presented in Bezruczko et al. from a classroom kinesiology achievement test. The construct in that research was not a psychological attribute; yet, it does not have properties of an additive extensive physical variable; hence, it presents comparable measurability issues. The study reported in Bezruczko et al. involved one hundred eight students enrolled in an undergraduate kinesiology class who took a 42-item classroom achievement test assessing the origin, insertion, and primary moving action of the shoulder musculature. The test consisted of 22 multiple-choice questions, 8 true/false questions, and 12 matching questions. All items were dichotomously scored.

The study included an achievement test model based on Bloom's qualitative taxonomy of educational objectives (Anderson, Krathwohl, and Bloom, 2000; Krathwohl, 2002), which are comprehension, memorization, application, evaluation, and analysis. Two of the co-authors independently matched each item with a Bloom educational objective. There was agreement on 38 items (90 percent). The two co-authors then discussed the four items for which originally there was no agreement and arrived at a classification for which there was agreement after discussion. The test contained 26 memorization items, six comprehension items, four evaluation items, three analysis items, and three application items.

Figure 1.
Marginal means of nodal centrality and item sequence order

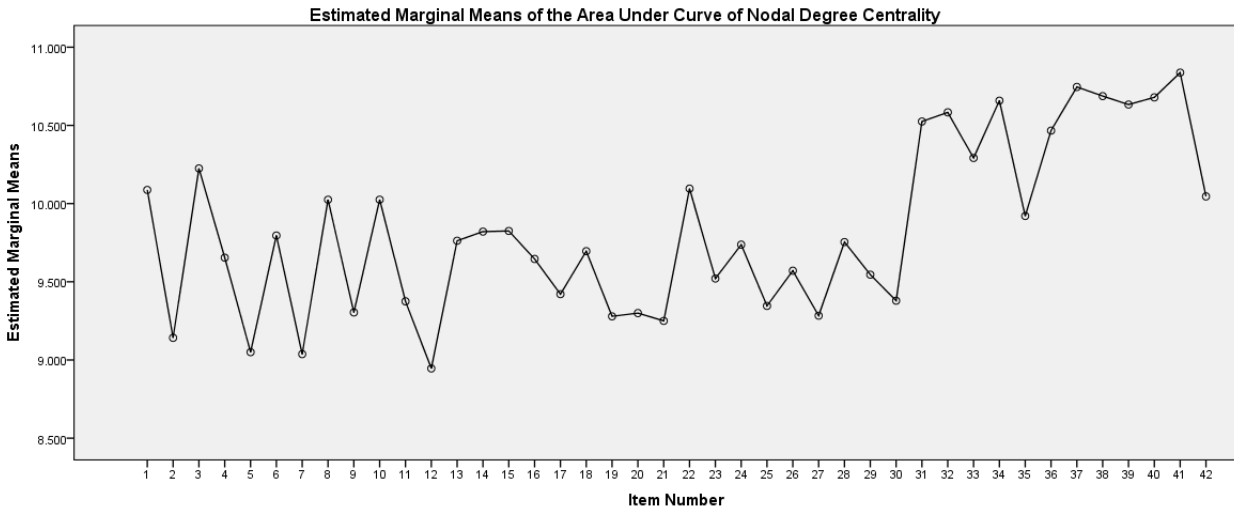
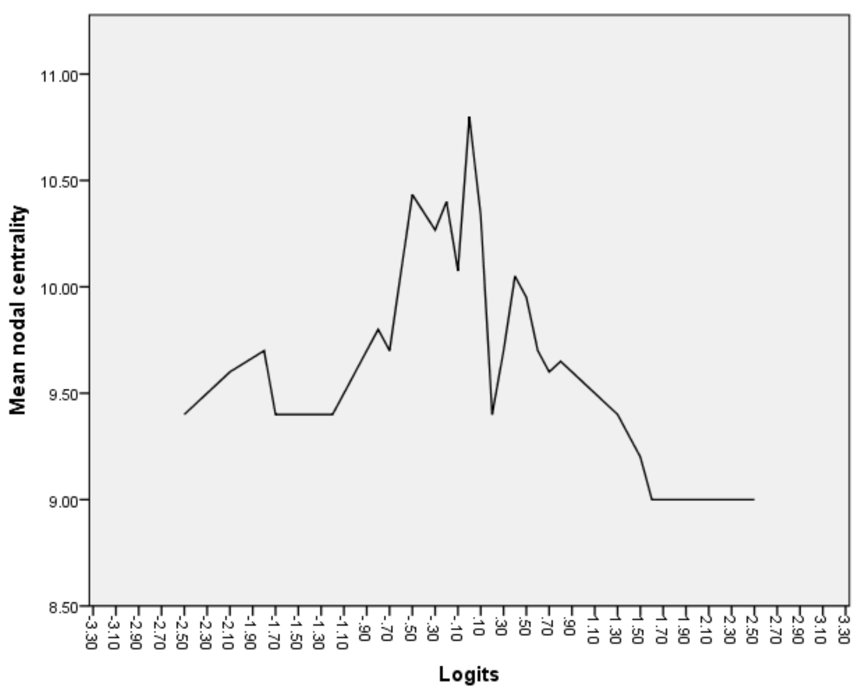


Figure 2.
Marginal means of nodal centrality and logit difficulty order



Bezruczko et al. hypothesized that these educational objectives represent a cognitive process dimension with the test items being ordered along this dimension. Specifically, it was hypothesized that memorization items would be easier than higher order cognitive operations and analysis and evaluation items should be more difficult than comprehension and application items. The results of the Rasch analysis confirmed that item content categories present a hierarchy that conforms to the hypothesized test model theory: "Comprehension and memorization items were significantly less difficult, while higher level cognitive operations such as application, evaluation, and analysis were progressively more difficult" (Bezruczko et al., 2016, p. 402).

While the present research is only in its early stages, preliminary analyses already show interesting results that we share with the RMT audience. Figure 1 presents marginal mean nodal centrality for items in sequence order, and our expectation is they should be uniform in a well-defined test, a continuous chain of linked nodes, which represent items. From a simple, Classical Test Theory perspective, these results seem to confirm that structure. Figure 1 shows a sequence of ordered items and a marginal nodal representation that is continuous with some irregularity but generally acceptable from an ordinal score perspective.

However, this structure does not address our question, which was about relations between a fundamental scale and graph theory results. Figure 2 presents results that dramatically differ when relations are examined between marginal nodal centrality and logit difficulties. While tails are generally uniform, mid-scale shows enormous "choppiness", which we speculate is related to content distribution. The middle of the scale (29 items) is dominated by a 2 to 1 mix of content, twice as many Memorization (19 items) versus everything else (10 items). The tails together have 13 items, seven Memorization versus everything else (six items). The overall test is 62 percent Memorization, the middle of the scale is 66 percent, while the tails are 54 percent. Our speculation is that this relatively balanced distribution is disturbed by an initial set of 12 Memorization items in mid-scale that shows virtually identical item difficulties, given published root mean square error = .36 for this test. That set of 12 items is functioning as a single item, which reduces Memorization in mid-scale distribution to 16 percent. In general, the choppiness of this nodal distribution does not support the underlying continuity of this ability despite adequate Rasch model item and person fits and reinforces concerns test developers frequently have about content representation.

These results are only preliminary, but they already point to interesting implications of graph theory. However, additional results appear to be pointing in direction of density of nodal connections as more accurate representation of continuity in an item network hence contributing to measurement validity. Planned analyses will examine density variation with item difficulty, and expectations are both high and low ability examinees will show systematic relations between nodal density and item difficulties.

References

Anderson, L. W., Krathwohl, D. R., and Bloom, B. S. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston:Allyn and Bacon.

Banasode, S. N. and Umathar, Y. M. (2018). Review of applications of graph theory in engineering. *International Journal of Mathematics Trends and Technology, 57*(4), 225-230.

Bezruczko, N., Frank, E., and Perkins, K. (2016). Does instruction affect the underlying dimensionality of a kinesiology test? *Journal of Applied Measurement, 17*(4), 393-415.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186.

Diestel, R. (2017). *Graph theory* (5th edition). Heidelberg: Springer-Verlag.

Edwards, E. S., Burke, K., Booth, J. R., and McNorgan, C. (2018). A complex network approach. *PloS one, 13*(12), e0208923.

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

Garner, M., and Engelhard, G. (2000). Rasch measurement theory, the method of paired comparisons, and graph theory, Vol. 5, 259-286. In M. Wilson and G. Engelhard Jr. (Eds.), *Objective measurement: Theory into practice*. Amsterdam: Elsevier.

Horowitz-Kraus, T. Buck, C. and Dorrman, D. (2016). Altered neural circuits accompany lower performance during narrative comprehension in children with reading difficulties: An fMRI study. *Annals of Dyslexia, 66*(3), 301-318.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*, 212-218.

Mao, J., Liu, L., Perkins, K. and Cao, F. Reading disability is characterized by a more connected network with wrong hubs. Under review.

Michell, J. (2003). Measurement: A beginner's guide. *Journal of Applied Measurement, 4*(4), 298-308.

Michell, J. (2004). Item response models, pathological science and the shape of error. Reply to Borsboom and Mellenbergh. *Theory & Psychology, 14*(1), 121-129.

Michell, J. (2005). The logic of measurement: A realist overview. *Measurement, 38*, 285-294.

Michell, J. (2013). Constructs, inferences and mental measurement. *New Ideas in Psychology,31*, 13-21.

Michell, J. (2020). The fashionable scientific fraud: Collingwood's critique of psychometrics. *History of the Human Sciences*, *33*(2), 3-21.

Piperski, A. (2015). An application of graph theory to linguistic complexity. *Yearbook of the Poznan Linguistic Meeting, 1*, 1, 89-102. DOI: 10.1515/yplm-2015-0005.

Resse, T. W. (1943). The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychological Monographs*, *55*(3), 6-20. Accessed October 29, 2020 at https://www.rasch.org/rn8.htm

Riaz, F. and Ali, K. M. (2011). Applications of graph theory in computer science. *IEEE Xplore*, DOI: 10.1109/CICSyn.2011.40.

Shirinivas, S. G., Vetrivel, S, and Elango, N. M. (2010). Applications of graph theory in computer science an overview. *International Journal of Engineering Science and Technology, 2*(9), 4610-4621.

Tatsuoka, M. M. (1986). Graph theory and its applications in educational research: A review and integration. *Review of Educational Research*, *56*(3), 291-329.

Wright, B. D., and Stone, M. H. (1979). *Best test design*. Chicago: MESA press.

Yang, J. and Tan, L. H. (2019). Whole-brain functional networks for phonological and orthographical processing in Chinese good and poor readers. *Frontiers in Psychology, 10,* 2945.

Statistics and Theory: A Public YouTube Channel

Vahid Aryadoust
National Institute of Education
Nanyang Technological University
Singapore
Email: vahid.aryadoust@nie.edu.sg

The Statistics and Theory is a public YouTube channel which was launched by Vahid Aryadoust of the National Institute of Education, Nanyang Technological University, Singapore. The channel has two main foci: quantitative data analysis and theoretical frameworks in (applied) linguistics and language assessment. The quantitative section includes, among other things, a series of videos on the application of Rasch measurements in assessments: https://www.youtube.com/playlist?list=PLTjlULGD9bNJqDb6omtmttAc1DIvaMMus The series also features Professor William Boone who presented on the use of the Rasch models for validating scales.

Currently, the Rasch measurement series features the use of dichotomous and polytomous Rasch models as well as many-facet Rasch measurement. In a near future, this collection will be expanded to include other extensions of Rasch measurement. Subscription and the use of the videos is free for everyone.

New normal? Let's learn new software!

Priyalatha Govindasamy [1]
Kathy E. Green[2]
Antonio Olmos[3]

[1]Universiti Pendidikan Sultan Idris, Malaysia
[2]University of Denver, USA,
[3]Aurora Research Institute, USA.

Work from home has been the new normal resulting from the CoVID-19 pandemic for many people around the globe. During the lockdown, many experienced difficulties in accessing statistical software because so many of them were licensed products and are installed on in-office computers. Therefore, accessibility to these software packages became scarce. In searching for solutions, we discovered that the statistical programming software R can address accessibility issues. R is a free open source and code software that is designed as a platform for scholars and practitioners for their research activities. The use of R is increasing gradually in all fields because it is very powerful, it is continuously updated and maintained by the top statisticians in the world, and it is free, open-source software. The R software requires different packages, just like other statistical programs require specific routines for specific purposes (i.e., modules in SPSS). The user can download those packages from the internet and install them, and it is recommended to run updates from time to time. There are a very large number of packages intended to run almost every imaginable analysis and the Rasch model is no exception.

There are altogether 50 packages available for Item Response Theory (IRT) analysis in the Comprehensive R Archive Network (CRAN) repositories in R. CRAN is the main repository for hosting R packages. However, only 16 packages were found focused on Rasch analysis. Among the Rasch packages, some focus entirely on the Rasch analysis but the rest comprise Rasch analysis as part of the package.

Table 1 lists the 16 packages designed for Rasch analysis. The similarities and differences between the packages are identified to organize the packages into clusters. The 16 packages can be grouped into 3 main clusters labelled as (1) model fit, (2) diagnostic, and (3) sample size estimation. The model fit cluster can be further grouped into regular and complex models. Under the regular model, four packages provide different estimation techniques in fitting the Rasch model to data.

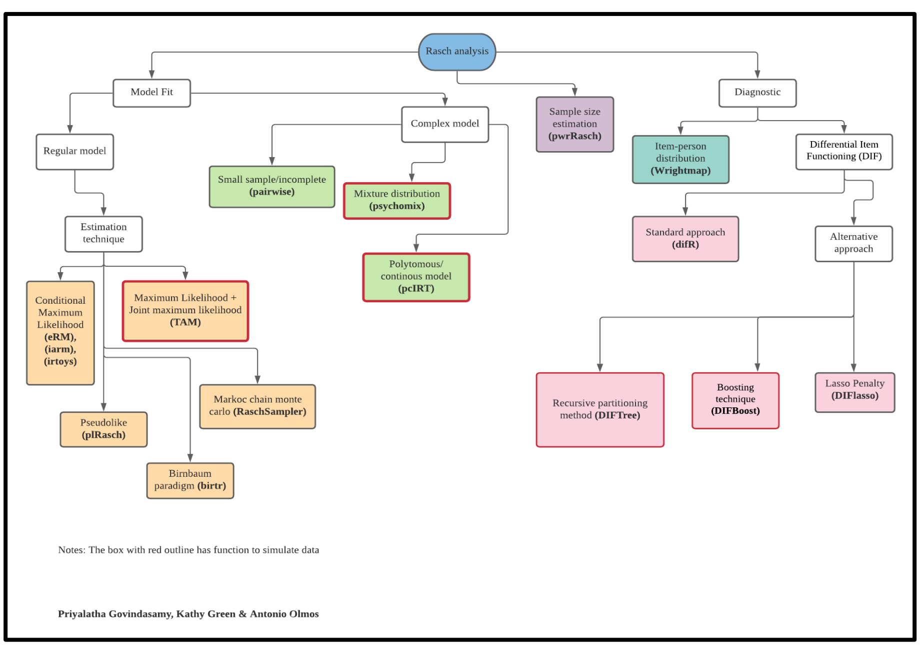
The **eRM**, **iarm**, and **irtoys** packages use Conditional Maximum Likelihood (CML) estimation while Markov Chain Monte Carlo estimation is used in the **RaschSampler** package. Maximum Likelihood (MLM) and Joint Maximum Likelihood (JML) are used in **TAM**. The Birnbaum paradigm estimation is used in the **birtr** package, and pseudolikelihood estimation is found in the **plRasch** package. Three packages were found to focus on the complex models of Rasch that account for non-dichotomous data. For small or incomplete data, the **pairwise** package is suggested. Then, the **pcIRT** package is designed for polytomous and continuous rating scale data. The

psychomix package is designed to capture the difference that exists within groups in a population.

Table 1.
List of Rasch Model packages in R

No	Package	Description
1	DIFTree	Item focused trees for the Identification of items in Differential Item Functioning
2	eRM	Extended Rasch Modeling
3	TAM	Test Analysis Modules
4	pcIRT	IRT models for Polytomous and Continuous Item Responses
5	Psychomix	Psychometric Mixture Models
6	difR	Collection of Methods to Detect Dichotomous Differential Item Functioning (DIF)
7	DIFlasso	A Penalty approach to Differential Item Functioning in Rasch Models
8	DIFboost	Detection Item Functioning (DIF) in Rasch Models by Boosting Techniques
9	plRasch	Log linear by Linear Association Models and Rasch family models by pseudolikelihood estimation
10	pairwise	Rasch Model Parameters by Pairwise Algorithm
11	RaschSampler	Rasch Sampler
12	pwrRasch	Statistical Power Simulation for Testing the Rasch Model
13	Iarm	Item analysis in Rasch Models
14	irtoys	A collection of functions related to Item Response Theory (IRT)
15	Wrightmap	IRT Item-Person Map with 'ConQuest' Integration
16	Birtr	The Basics of Item Response Theory using R

Figure 1.
Illustration of package grouping



The second cluster named Diagnostic can be further explained by two groups called item-person distribution and differential item functioning (DIF). It is labelled as a diagnostic cluster because these packages are designed specifically to evaluate the performance of items and persons after fitting a Rasch model. The **Wrightmap** package allows us to map the item difficulty parameter by proficiency distribution of respondents. This visual representation indicates any problematic items or persons targeted for further actions. Four other packages assess differentially functioning items. The **difR** and packages use a standard approach such as logistic regression, Mantel-Haenszel, Standardization, Lord's chi-square in evaluating the item functioning within the sample. The **DIFboost**, **DIFTree**, and **DIFlasso** packages propose alternative approaches in estimating differential item functioning in the data. The DIFboost package uses a boosting technique in detecting dif whereas the DIFlasso package uses the group lasso penalty approach when detecting differential item functioning. The DIFTree capture differentially functioning items using recursive partitioning method (trees).

Finally, the third cluster is labelled as sample size estimation. This cluster consists of a single package called **pwrRasch**, which helps to determine the appropriate sample size for Rasch analysis. Additionally, five packages (TAM, psychomix, pcIRT, DIFTree, and DIFBoost) across the clusters have a common function allowing us to simulate data. The DIFTree and DIFboost packages offer data simulation that comprise differentially functioning items. Data to fit a standard Rasch can be simulated using the TAM package. In pcIRT, data can be simulated for the Multidimensional polytomous Rasch model (MPRM) and Continuous Rating Scale model (CRSM). The psychomix package provides the opportunity to generate data for Mixture Rasch Model. Figure 1 presents a summary of the package grouping. The programming nature of the R software could be overwhelming for first-timers. However, there are many free, readily available tutorials and information for users of different R proficiency levels to master the software. The *R for Dummies* book could be a good start for beginners to learn the software.

Call for Nominations: Georg William Rasch Early Career Publication Award

We are pleased to announce that we are currently accepting nominations for the Georg William Rasch Early Career Publication Award, which is an AERA-sanctioned award. This award is presented to an individual for outstanding Rasch measurement research published within five years of obtaining their doctoral degree and will be presented during the AERA 2021 Annual Conference. The main purpose of this award is to foster ongoing quality research in the area of Rasch measurement, and to encourage the development of a Rasch measurement focus in the early phases of one's career.

Eligibility Criteria: Nomination for the Georg William Rasch Measurement Early Career Publication award should be based on a scholarly publication authored by a nominee that fulfills the following criteria:

- The publication must include a Rasch measurement focus.
- The publication may be based on the dissertation work of the nominee or other recent research the nominee has conducted.
- The nominee should be either the single author or the lead author (in the case of a jointly authored paper) of the article.
- The article must have been published within two (2) calendar years (April 2019 – April 2020) prior to the Rasch Measurement SIG's business session at the 2021 AERA Conference.
- Only peer-reviewed research manuscripts that have been published or are in press (accepted for publication) are eligible for nomination.
- The nominee should have received his/her doctoral degree no earlier than five years prior to the nomination deadline (January 1, 2016).
- The nominee must be a member of the Rasch SIG or must become a member by the time the award is presented at the annual meeting.

The Award: The award includes a monetary stipend and a plaque that includes the name of the award, the winner's name, the title of the winning article, and the name of the journal or peer reviewed research publication in which the article was published. The award will be given to one person, biannually in odd-numbered years.

Nomination Deadline: The nomination deadline is Friday, January 1, 2021.

How To Submit: Nominations are to be submitted by sending an email to the convener of the Awards Committee Dr. Eli Jones, eli.jones@memphis.edu. Completed nominations should include ALL the following:

- A letter nominating the author of an early career publication. Please include the name of the author, the date he/she received the doctoral degree, and the name of the institution that conferred the degree. The nominator's letter must provide reasons that the paper is an example of an outstanding Rasch measurement research publication.
- A copy of the published paper, including complete bibliographic information.
- A copy of the Table of Contents of the journal or other peer reviewed research publication in which the paper appeared OR an acceptance letter from the journal will be acceptable if it is currently in press.
- A current CV for the nominee.

Please visit our website for the latest news and announcements: https://www.aera.net/SIG083/News-Announcements

List of Recent Publications in Journal of Applied Measurement

Vol. 21, No. 3, Fall 2020

A-priori Weighting of Items with the Rasch Model
David Andrich and Sonia Sappl

The Equivalence of the Test Response Function to the Maximum Likelihood Ability Estimate for the Dichotomous Rasch Model: A Proof
Ben Babcock

Many-facet Dichotomous Rasch Model Analysis of the Modern Language Aptitude Test
Mitch Porter and Stefanie A. Wind
How Well Do Simulation Studies Inform Decisions About Multistage Testing?
Wenhao Wang, Jie Chen, and Neal Kingston

Examining the Pre-service School Principals' Impromptu Speech Skills with a Many-Facet Rasch Model
Mingchuan Hsieh and Akihito Kamata

Evaluating Longitudinal Anchoring Methods for Rasch Models
Tara L. Valladares and Karen M. Schmidt

Exploring the Psychometric Properties of a Self-Efficacy Scale for High School Students
Yuan Ge and Stefanie A. Wind

Assessing Differential Statement Functioning in Polytomous Multidimensional Pairwise Comparison Items
Xue-Lan Qiu

Response Differences in Appraisals of Working Conditions among Elementary and High School Teachers
Richard G. Lambert, C. Missy Moore, Christopher McCarthy, and Bryndle L. Bottoms

Upcoming Rasch Measurement Courses and Workshops

The following Rasch-related courses will be offered via statistics.com in the coming months.

Rasch Measurement: Core Topics
Instructor: Everett V. Smith, Jr
Dates: January 22, 2021 to February 19, 2021 and May 21, 2021 to June 18, 2021
For more details, please see:
https://www.statistics.com/courses/rasch-measurement-core-topics/

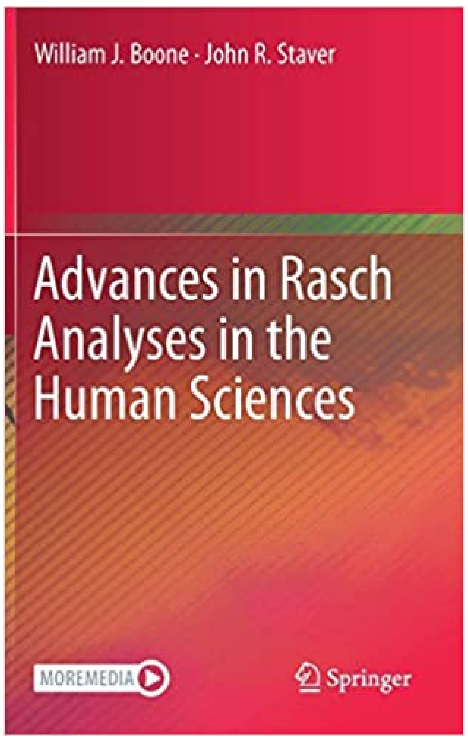
Rasch Measurement: Further Topics
Instructor: Everett V. Smith, Jr
Dates: June 25, 2021 to July 23, 2021 and June 24, 2022 to July 22, 2022
For more details, please see:
https://www.statistics.com/courses/rasch-measurement-further-topics/

Many-Facet Rasch Measurement
Instructor: Everett V. Smith, Jr
Dates: August 13, 2021 to September 10, 2021 and August 12, 2022 to September 9, 2022

For more details, please see: https://www.statistics.com/courses/rasch-measurement-many-facet/

Book Announcement: Advances in Rasch Analyses in the Human Sciences

William J. Boone & John R. Staver



We are pleased to announce the publication of *Advances in Rasch Analyses in the Human Sciences*.

Key features of this text are as follows:

- Presents an easy to digest manner of complicated material
- Offers numerous activities to practice the reader's understanding of Rasch techniques
- Provides links between the chapters of this and the previous volume
- Includes model write-ups that can be used to explain the chapter topic in articles and presentations
- Offers step-by-step instruction on how-to use Rasch analysis
- Features a combination of Rasch conversations, Rasch data analyses, Rasch theory, and formative assessments
- Written for a broad audience as an introductory text: light on the mathematics, heavy on the practical aspects, and clear on the concepts and theory
- All activities and tables utilize user-friendly (and free) Ministeps Rasch software that prepares readers to confidently utilize Rasch Winsteps software

Interested readers can find out more about this book at the following webpage: https://www.springer.com/gp/book/9783030434199