Fall 2018, VOL. 32 NO. 1; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

Notes:

Transactions of the Rasch Measurement SIG
American Educational Research Association

Note from the New Editors

Starting with this first issue of the 32nd volume of the Rasch Measurement Transactions (RMT), the RMT editorship has changed hands to two new co-editors: Leigh Harrell-Williams and Stefanie A. Wind. We want to thank Ken Royal for his years of service to the Rasch community as RMT editor. Ken took over the reins from Mike Linacre in 2012. We'd also like to thank Mike Peabody for his ongoing work as the Rasch.org webmaster.

The first volume of RMT appeared in 1987, and we'd like to keep the publication as relevant to the Rasch community today as it was in the early years. More and more journals are publishing Rasch measurement work, which is good for researchers and practitioners. However, contributions to RMT have sometimes waned. We will be making a few changes to encourage submissions and ensure a variety of applications and methodological issues are highlighted.

For the next few editions, we will be showcasing the work of up-and-comers in the Rasch community that was presented at International Objective Measurement Workshop (IOMW) held in April 2017, prior to AERA and NCME. The IOMW organizers, Andy Maul and Ronli Diakow, shook things up by having 6 short talks in the opening plenary instead of one speaker. We'd like to keep that momentum going by highlighting the work of these presenters. (We should disclose that both Stefanie and Leigh were part of this session).

Rasch Measurement Transactions
www.rasch.org/rmt
Copyright © 2018 Rasch Measurement SIG, AERA
Permission to copy is granted.

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

We'll also be asking for those of you who have recently presented at non-measurement conferences to write a short summary about your talk so we can highlight your work.

Finally, we'll be working with our webmaster to figure out the logistics of a submission form on the Rasch.org website. We'd like to know what you want and need from RMT, so we'll be sending out a survey sometime soon to get feedback on our thoughts and solicit your ideas.

Sincerely,

Your new RMT Co-editors, Leigh and Stefanie

Note from Chair of the Rasch Measurement SIG

Dear Rasch SIG Colleagues,

Greetings! It is my great honor to serve as Chair of the Rasch SIG for 2018 to 2020!
First, I would like to introduce the new officers to you:

- Cari F. Herrmann-Abell, Secretary, *American Association for the Advancement of Science*
- E. Matthew Schulz, Treasurer, *Smarter Balanced Assessment Consortium*
- Eli Andrew Jones, Program Chair, Columbus State University
- Trent Haines, Co-Program Chair, Morgan State University

At different stages of my professional development, I have benefited greatly from researchers devoted to Rasch measurement. I know a number of colleagues in the Rasch SIG have had the same experience. Thus, I would like to encourage you to invite colleagues, collaborators, and graduate students in your institution and elsewhere to join the Rasch SIG. We are a welcoming and collegial organization and we all hope to see greater involvement from each of you.

For the current Rasch SIG members, it is time to renew your AERA membership. Please make sure that you include the Rasch SIG membership with additional dues of $10.00 in your renewal to continue your involvement with us. When you go to the AERA website, if "Renew with changes" is chosen, the Rasch SIG box may not be checked. If not, your Rasch SIG membership will not be renewed automatically. Thus, please double check before you proceed to checkout that you have the Rasch SIG membership included.

As the past Chair Dr. Leigh Harrell-Williams indicated that membership in the Rasch SIG is important. The Rasch SIG presents a platform to allow researchers, either methodologists or applied researchers to share their Rasch related work at AERA. Further, the membership fees are used to support the Rasch SIG awards, maintain the Rasch SIG website (raschsig.org), and cover the costs for the annual business meeting. Your support for the Rasch SIG is needed, valued, and appreciated. That support is used to encourage and acknowledge the contributions of researchers to Rasch measurement. I will be reaching out to ask each of you to join us again, based on the information provided by AERA, to contact members who have not renewed in the past two years. Our membership determines the level of involvement we can have in the AERA national meetings, among other matters.

As I mentioned above, the Rasch SIG gives two awards: The Georg William Rasch Early Career Publication Award and the Benjamin Drake Wright Senior Scholar Award. I would like to encourage you to think about potential nominations for these awards. More details regarding the exact nomination process and deadline (January 2019) will be sent to all our SIG members via email.

I would like to invite the Rasch SIG members to reach out to me with questions, concerns or suggestions regarding the future. This includes how to improve promotion of the Rasch SIG. If you have any publications such as books, papers, or new software, or are willing to share course syllabi or other resources, please contact me or Leigh so that such information can be updated onto the Rasch website. My email is hjiao@umd.edu and Leigh's email is leigh.williams@memphis.edu.

Looking forward to hearing from you and serving the Rasch SIG over the next two years.

Hong Jiao
Rasch SIG Chair, 2018-2020

Biographies for the New Rasch SIG Officers

Chair: *Hong Jiao*

Hong Jiao is an associate professor at the University of Maryland (UMD), College Park specializing in educational measurement and psychometrics in large-scale assessment. Prior to joining the faculty in Measurement, Statistics, and Evaluation at UMD, she has worked as a psychometrician at Harcourt Assessment. Dr. Jiao has served on different national and international professional associations. As Director of Maryland Assessment Research Center (MARC), she works with the team to provide psychometric research and service to the State assessment programs. She co-organized recent MARC annual conferences on different cutting-edge topics in assessment including technology-enhanced innovative assessment, data analytics and psychometrics, applications of artificial intelligence in assessment and co-edited the books on these topics. She has published and presented on a variety of topics, including multilevel item response theory

modeling, modeling complex local item dependence in innovative assessments, mixture item response theory modeling, integrating item responses and response time for cognitive diagnosis, and psychometrics in large-scale assessment. She is currently co-editing a special topic in Frontiers in Psychology on Process Data in Educational and Psychological Measurement. A lot of her work is related to Rasch modeling and the extensions of the Rasch model.

Secretary: *Cari F. Herrmann-Abell*

I am a Senior Research Associate at AAAS Project 2061. I received my Ph.D. in Chemistry from the University of North Carolina. My current research focuses on the development and use of assessments of physical science topics. I am currently a principal investigator on measurement grant funded by the U.S. Dept. of Education's Institute of Education Sciences. The goal of the project is to develop NGSS-aligned three-dimensional assessment tasks that can be used to assess students understanding of energy ideas. Rasch has played a major role in our item and test development process and in our evaluation of the curriculum materials we develop. In addition, I also lead workshops on the item development process for researchers and classroom teachers. During the workshops, one of my goals is to increase practitioners' awareness and understanding of Rasch and the role it can play in understanding what students know and do not know.

Treasurer: E. Matthew Schulz

Matt Schulz is an 1987 graduate of the University of Chicago. His dissertation on applications of Rasch Measurement in rehabilitative medicine was guided by Benjamin Wright. Dr. Schulz worked briefly as Coordinator of Measurement, Statistics and Evaluation at Chicago Public Schools, Director of Testing at the National Council of State Boards of Nursing, and then settled into a seventeen year engagement in the statistical research department at ACT. This was followed by nine years at Pacific Metrics Corporation where he led an online, state-wide, end-of-course testing program and later served as vice president of research and test development. Dr. Schulz is currently Director of Psychometrics at the Smarter Balanced testing consortium. He continues to work intermittently as a consultant and technical adviser to testing companies, state departments of education, and other organizations. Dr. Schulz made early and frequent contributions to the development of computer programs for Rasch measurement such as BIGSTEPS and MSTEPS, to the Rasch SIG Newsletter, and has authored chapters in various edited volumes on Rasch measurement. He has published extensively on measurement topics in a wide variety of journals including JEM, JAP, APM, and JEBS.

Program Co-Chair: *Eli Jones*

Eli Jones is assistant professor of research at Columbus State University. He received his doctorate in Educational Inquiry, Measurement, and Evaluation from Brigham Young University, and completed postdoctoral work in teacher evaluation and psychometrics at the University of Missouri's Network for Educator Effectiveness. His published studies have been included in *Educational and Psychological Measurement*, the *Journal of Educational Measurement, Educational Assessment,* the *Journal of Applied Measurement,* and the *Journal of Research on Educational Effectiveness.* His research interests include performance assessments, rater-mediated assessments, teacher evaluation, K-12 academic interventions, and applications of the Rasch model. He is particularly interested in methodological issues related to educational assessment including rater error, response patterns, and rating designs.

Program Co-Chair: *Trent Haines*

R. Trent Haines is an Associate Professor at Morgan State University where he teaches in and serves as Program Coordinator for the Graduate Program in Psychometrics. His primary area of research interest is in the use of Rasch measurement models to develop and evaluate culturally responsive measures.

Call for Nominations: Georg William Rasch Early Career Publication Award

I am writing to encourage you to submit an award nomination for The Georg William Rasch Early Career Publication Award, which is an AERA-sanctioned award. The award shall be presented to an individual for outstanding Rasch measurement research published within five years of obtaining their doctoral degree. This is the 3rd time that we'll be giving this award.

The award is being offered as an incentive to foster ongoing excellent research in the area of Rasch measurement through the early phases of one's career.

Eligibility Criteria:

- The publication (with a Rasch measurement focus) being nominated for the Georg William Rasch Measurement Early Career Publication Award 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 an article published in the 12 months of the calendar year prior to the Rasch Measurement SIG's business session at the annual AERA meeting.

- Only peer-reviewed research publications are eligible for nomination.

- The nominee should have received his/her doctoral degree no earlier than five years prior to the nomination deadline.

The Award

The award includes a stipend and a plaque that includes the name of the award (The Georg William Rasch Measurement Early Career Publication 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.

The deadline for nominations is Friday 15 January 2019. Nominations are submitted by sending an email to the convener of the Awards Committee proposing the name of the nominee and describing the grounds on which the nominee meets the requirements for the award.

Nominations should include (and are restricted to) 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 include 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

- A current CV for the nominee

Cari Herrmman-Abell
Convener of Awards Committee
carifha@gmail.com

Recognizing George Engelhard, Jr.: Recipient of the 2018 Benjamin Drake Wright Senior Scholar Award

George Engelhard, Jr. was the 2018 recipient of the Rasch Measurement Theory Special Interest Group's *Benjamin Drake Wright Senior Scholar Award*. Professor Engelhard was a student of Professor Wright at The University of Chicago, and his program of research reflects Ben's commitment to the improvement of measurement practice through careful thinking and clear communication.

Professor Engelhard's program of research is focused on the quest for invariant measurement within a variety of contexts in the human sciences. His work includes methodological contributions, historical and philosophical explorations, and applied research related to Rasch measurement theory. In addition to general methodological work related to the Rasch model, Professor Engelhard's research includes the application of Rasch measurement theory to the area of rater-mediated assessments. In this area, he focuses on the development of Rasch-based indicators that can be used to evaluate the quality of rater-mediated assessments and inform the development and revision of operational assessment systems that involve raters.

Professor Engelhard's work reflects his commitment to the fundamental requirements for invariant measurement as a perquisite to the interpretation of measurement outcomes is clear. His commitment to the improvement of education is demonstrated through his research contributions and international renown as a dedicated mentor who is eager to share his findings and encourage sound measurement practices across a variety of contexts.

Stemming from this program of research is a plethora of noteworthy publications that have had a significant impact on measurement practice and pedagogy across international contexts. In particular, Professor Engelhard's publications are known for their didactic nature and careful consideration of implications for research, theory, policy, and practice. In his research, Professor Engelhard's focuses on examining practical measurement issues from the perspective of invariant measurement in order to inform research, theory, policy, and practice across a variety of contexts in the human sciences. Although the substantive areas in which his work has appeared are diverse, each of his contributions is characterized by a clear emphasis on the fundamental requirements for measurement. Professor Engelhard's work reflects the mission of the Rasch measurement community, and it brings attention to the critical importance of developing empirically testable instruments with objective measures using strong measurement frameworks.

Evidenced by his many co-authored publications, Professor Engelhard is a dedicated teacher and mentor. Many scholars have been significantly influenced by Professor Engelhard's generous mentorship. Beyond these collaborative relationships, Professor Engelhard has had a major influence on measurement practice in the United States and international contexts through his service on technical advisory committees and workshops across the world.

Below is the abstract that George provided for his talk at the Rasch SIG Business meeting. The slides from the presentation are found on the Rasch.org website: https://www.rasch.org/rmt/rmt321_keynote.pdf

Stefanie A. Wind
The University of Alabama

Power, Invariance, and Measurement
George Engelhard Jr.
Rasch Measurement SIG Meeting Presentation

Measurement serves numerous purposes in the social sciences. For example, test scores in education are used to guide decisions (diagnostic, formative, summative and strategic) that are made about students. In this presentation, I focus on measurement as a powerful technology that has great potential for both positive and negative consequences in education. Next, I describe the principles of invariant measurement, and their use to guide the construction of stable and useful assessment systems. Invariant measurement refers to assessments that yield test scores that "stay put while the user's back is turned". Finally, I introduce Rasch measurement theory as a psychometric model for creating stable systems of educational assessments. Rasch models offer an approach for developing invariant measures that can support the intended uses of educational assessments. This presentation provides an overview of these concepts, as well as the implications for research, theory and practice related to educational assessments.

Best Test Design and _Rating Scale Analysis_ are now available for free download!

Download the pdfs at:
https://rasch.org

Summary of Rasch SIG Business Meeting, April 2018

Outgoing and incoming officers were introduced. Membership numbers and finances were reviewed. Given our current rate of spending, we will deplete our funds soon. We spent much of the meeting discussing ways to improve the financial outlook of the SIG. One possible suggestion was to reduce the award amount and/or eliminate the travel award portion of the senior researcher award. This action would require a change to the SIG bylaws. Some members suggested that the travel award would be more appropriate to offer to the early career awardee instead of the senior awardee. Most members agreed that the award amounts should be decreased. Another suggestion was to give the awardee the option to waive the travel award.

We also discussed changing the time and location of the business meeting to encourage more attendance. Another possibility was to join forces with another SIG or move the meeting so that IOMW attendees could also attend the SIG meeting possibly to a nearby restaurant.

The members also agreed that the SIG needs to do a better job of advertising the senior scholar talk. Finally, we discussed the possibility of offering a pre-session at NCME or AERA as a way of building membership.

Cari Herrmman-Abell

Upcoming Course on Rasch Measurement

In addition to the online courses that continue to be provided, and in collaboration with The University of Sydney, a course in Rasch Measurement Theory will take place over the period 11 & 12 and 15 - 19 July 2019, in Perth, Australia at The University of Western Australia (UWA).
Further information is available at http://www.education.uwa.edu.au/ppl/courses and the administrative contact is natalie.carmody@uwa.edu.au.

IOMW 2018 Recap

The 2018 meeting of the International Objective Measurement Workshop (IOMW) was held on three chilly days in April at the Kimmel Center of New York University, next to the lovely Washington Square Park. Thanks to the generosity of IOMW's sponsors – Modus Outcomes, MetaMetrics, Educational Data Systems, and the Center for Practice and Research at the Intersection of Information, Society, and Methodology (PRIISM) at NYU – registration fees were even lower than they had been in previous years, despite being in a more expensive location; possibly as a result, attendance was about 20% higher than it was at the previous meeting. This made for a vibrant and engaging conference, with a mixture of new and familiar faces and voices.

The emergence of a new generation of scholars working on vital measurement issues was apparent throughout the conference. The morning of April 10 kicked off with a series of presentations by young scholars, including graduate students and professors in the first decade of their career. Their presentations highlighted a range of new applications and theoretical developments, and gave a sense that the spirit of IOMWs past – in particular, the willingness to rigorously and thoughtfully work through measurement issues of vital relevance to many aspects of human activity – is not only alive and well, but is continuing to mature and take on new dimensions as society itself evolves.

The dialogue continued during a lunch poster session and the afternoon paper sessions, which addressed a wide range of topics including matters mathematical, philosophical, metrological, and economical. (Not vegetable or mineral, though, as far as I saw.) In the evening, attendees mingled at the Carroll Place, enjoying free food and drink (thanks, again, to the generosity of our sponsors!) and the chance to catch up with and meet colleagues old and new.

Inter-generational dialogue was further facilitated by a new mentoring program, spearheaded by Stefanie Wind, which paired established and newer scholars in a more formal way. Mentors read and provided feedback on their mentees' papers in advance of the conference, and then met in person at the conference for further discussion. Based on the enthusiastically positive feedback, future meetings of IOMW will likely continue some version of this program.

Wednesday morning kicked off with a talk by invited speaker Bob Mislevy, who explored how sociocognitive perspectives can enrich our thinking about measurement. Paper sessions that followed again explored a rich range of topics, including a review of a book celebrating the career and life of Ben Wright.

The conference concluded with a brief discussion of the future of IOMW. The next meeting will be held in Berkeley in 2020, led by IOMW veterans and rising stars Veronica Santelices, Yukie Toyama, and Perman Gochyyev. During this discussion it was noted that the name itself might deserve some rethinking: first, because the word "workshop" might not be as appropriate a label now as it was when IOMW was first created, and second, because the phrase "objective measurement" is increasingly recognized as being redundant. The new IOMW organizational committee may have more to say on this issue soon.

The ever-popular IOMW software workshops were held on April 12, spearheaded by Mark Moulton. In the morning, presentations were given on RUMM, Winsteps, Facets, ConQuest, Damon, jMetrik, and OpenBUGS. A panel discussion on big data took place over a catered lunch, followed by afternoon breakout sessions on individual software packages.

I'd like to extend my own sincere thanks to Ronli Diakow and the rest of the IOMW 2018 Organizational Committee for everything they did to make the meeting a success, as well as to all the attendees for their excellent presentations and energetic participation in discussions. IOMW has meant a lot to me in my own development as a scholar over the years, and I feel honored to be able to witness its continued

evolution as a gathering place for an increasingly diverse group of scholars, united by a willingness to think deeply and critically about measurement theory and practice.

Andrew Maul
University of California, Santa Barbara

Journal of Applied Measurement
Vol. 19, No. 1, Spring 2018

The Impact of Missing Values and Single Imputation upon Rasch Analysis Outcomes: A Simulation Study - *Carolina Saskia Fellinghauer, Birgit Prodinger, and Alan Tennant*

Methods for the Comparison of Differential Item Functioning across Assessments -*W. Holmes Finch, Maria Hernández Finch, Brian F. French, David E. McIntosh, and Lauren Moss*

Equating Errors and Scale Drift in Linked-Chain IRT Equating with Mixed-Format Tests - *Bo Hu*

Validation of Response Similarity Analysis for the Detection of Academic Cheating: An Experimental Study - *Georgios D. Sideridis and Cengiz Zopluoglu*

Rasch Analysis of the Teachers' Knowledge and Use of Data and Assessment (tKUDA) Measure - *Courtney Donovan*

Psychometric Properties and Differential Item Functioning of a Web-Based Assessment of Children's Social Perspective-Taking - *Beyza Aksu Dunya, Clark McKown, and Everett V. Smith*

Assessment of Test Items with Rasch Measurement Model - *Patrick U. Osadebe*

Richard Smith, Editor, www.jampress.org

Journal of Applied Measurement
Vol. 19, No. 2, Summer 2018

Comparing Disability Levels for Community-dwelling Adults in the United States and the Republic of Korea using the Rasch Model - *Ickpyo Hong, Annie N. Simpson, Kit N. Simpson, Sandra S. Brotherton, and Craig A. Velozo*

Using the Rasch Model to Investigate Inter-Board Comparability of Examination Standards in GCSE - *Qingping He and Michelle Meadows*

Using Repeated Ratings to Improve Measurement Precision in Incomplete Rating Designs - *Eli Jones and Stefanie A. Wind*

The Impact of Differential Item Functioning on the Warwick-Edinburgh Mental Well-Being Scale - *Hong Eng Goh, Ida Marais, and Michael Ireland*

Rasch Analysis of the Brief Symptom Inventory-18 with African Americans - *Ruth Chu-Lien Chao, Kathy Green, Kranti Dugar, and Joseph Longo*

Development and Calibration of Chemistry Items to Create an Item Bank, using the Rasch Measurement Model - *Joseph N. Njiru and Joseph T. Romanoski*

Psychometric Evaluation of the Revised Current Statistics Self-efficacy (CSSE-26) in a Graduate Student Population using Rasch Analysis - *Pei-Chin Lu, Samantha Estrada, and Steven Pulos*

Richard Smith, Editor, www.jampress.org

Featured IOMW Plenary Session Paper Summary:

Item Difficulty Modeling of Reading Comprehension Items Using the Rasch Latent Regression Linear Logistic Test Models

Introduction

The Common Core State Standards call for students to read increasingly complex texts as they progress through grades K-12 (National Governors Associate Center for Best Practices [NGACBP] & Council of Chief State School Officers [CCSSO], 2010). Although a few studies used psychometric models to identify sources of difficulty in reading comprehension (RC) with adult readers (e.g., Embretson & Wetzel, 1987; Gorin & Embretson, 2006), such studies are surprisingly sparse with a K-12 sample. This study used explanatory item-response models (De Boeck & Wilson, 2004) to investigate the effects of passage- and item-features of RC items from a computer-adaptive test (CAT) on item difficulty across a wide range of developmental levels. Specifically, the study examined which sets of predictors – ones related to passage processing (e.g., average sentence length, mean log word frequency) or those related to processing items and selecting an answer choice (e.g., falsifiability of answer choices) – explained more of variance in item difficulty.

Methods

The data came from a placement test into an online intervention program called ReadingPlus®. The assessment was given as computer-adaptive testlets, each composed of a passage (about 170-250 words in length) and five multiple-choice questions. Each passage was designed to represent one of 12 levels of complexity, corresponding roughly to 12 grade levels in terms of vocabulary demand. An analytic sample included 10,547 students in grades 1-12 in North America. This sample was randomly split into two subsets for model building and cross validation.

The study used the Rasch Latent Regression Linear Logistic Test Model (RLR-LLTM, Wilson & De Boeck, 2004), a doubly explanatory IRT model, which extends the Rasch Model (1) on both the item side (2) and the person side (3):

$$logit(\Pr(Y_{ip} = 1 \mid \theta_p) = \theta_p - \delta_i \quad (1)$$

$$\delta_i = \sum_{k=0}^{K} \beta_k X_{ki} \quad (2)$$

$$\theta_p = \gamma Vocab_p + \mu_p \quad (3)$$

where θ_p is reading ability for person *p*, δ_i is difficulty of item *i*, X_{ki} is item *i*'s score for item feature *k*, and β_k is the effect of item feature *k* on δ_i. Figure 3 shows the list of models proposed in the literature for text processing (called text representation models, Embretson & Wentzel, 1987) as well as item/item-passage predictors affecting response decisions used as predictors (X_{ki}).

Student vocabulary level ($Vocab_p$) was included to account for between-person differences in reading ability and to obtain more precise estimation of item parameters because it was used to select the first testlet (Mislevy & Sheehan, 1989; Mislevy, 1987). All the three equations were concurrently run in a single analysis, increasing better parameter estimates with smaller measurement errors. We note that the explanatory variables were all standardized so that regression coefficients would be comparable.

Models were compared primarily by pseudo-R^2 proposed by Embretson (1983, Figure 1) as well as by standard model fit criteria, the Akaike information criteria (AIC, Akaike, 1974) and Bayesian information criteria (BIC, Schwarz, 1978).

$$\Delta^2 = \frac{\ln L_0 - \ln L_m}{\ln L_0 - \ln L_S}$$

where:
lnL_0 = *log-likelihood for the null model with just constant value for item difficulty [Eq. (1) +(3) + (2) with only intercept* β_{0i} *No* X_{ki} *]*
lnL_S = *log-likelihood for the saturated model [Eq. (1) +(3)]*
lnL_m = *log-likelihood for the model to be evaluated [Eq. (1) + (2) + (3)]*

Figure 1. Pseudo-R^2 / fit index (Embretson, 1983) used to evaluate models

A. Text Representation Models		
Gorin-Embretson (2006)	• modifier propositional density (count of adjectives per 1000 words)	• predicate propositional density (count of verbs per 1000 words) • mean log word frequency
Lexile	• mean log word frequency	• mean sentence length
Coh-Metrix	• narrativity • syntactic simplicity • word concreteness • referential cohesion	• deep/causal cohesion • verb cohesion • connectivity • temporality
Text-Evaluator	• academic vocabulary • word unfamiliarity • concreteness • syntactic complexity	• lexical cohesion • level of argumentation • degree of narrativity • interactive/conversational style
B. Response Decision Predictors		
Item-type	Questions are classified into one of the four levels, ranging from 1) text-base: the answer is explicitly stated within a single sentence in verbatim) to 4) knowledge-base: the answer is not explicitly stated in the source passage and the reader has to infer using background knowledge (Ozuru et al., 2008).	
Abstractness of information requested by the question	Questions are classified into one of the four levels, ranging from 1) requesting highly concrete information to 4) highly abstract information (Ozuru et al., 2008, based on Monsenthal, 1996)	
Falsifiability	The number of falsifiable distractors (a possible value ranged from 0 to 3) (Embretson & Wentzel, 1987; Ozuru et al., 2008)	
Vocabulary demand of: - question - correct answer - distractors	Flesch-Kincaid grade level (a continuous variable ranging from 0 to 12, with 12 as the highest level) was calculated for: the question, the correct answer, and distractors (Embretson & Wentzel, 1987; Gorin & Embretson, 2006)	

Figure 2. List of passage- / item-predictors used in the study

Results

First text representation (TR) models listed in Panel A of Figure 2 were run, followed by models with response-decision (RD) predictors listed in Figure 2 Panel B. The great majority of variance in item difficulty was explained by the TR models (pseudo R^2 values ranged from 0.45 for the Gorin & Embretson model to 0.56 for the Lexile and Coh-Metrix with RD predictors show limited explanatory power (pseudo R^2 value ranged from 0.02 to 0.26). Indeed, when the best response model was combined with the best text-representation model, the amount of additional variance increased only by 3% points.

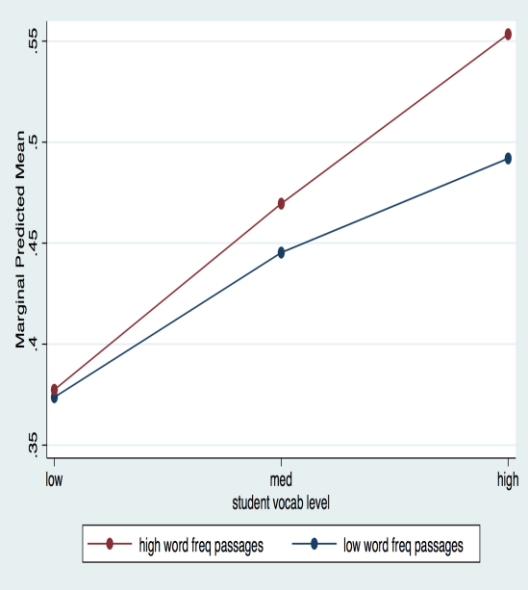
In terms of specific predictors, the average sentence length and the log mean word frequency, the traditional two predictors of readability, used in the Lexile, had relatively large effect on item difficulty (0.24 and -0.11 respectively) according to the TR and RD combined model (Figure 3).

Additionally, the use of temporal connectives (e.g., first, until) had a small but statistically significant negative effect on item difficulty while the use of adversative/contrastive connectives (e.g., although) had a small, statistically positive effect. (However, the connectivity was not found a significant predictor in the cross-validation sample). Surprisingly, items that require rewording/reconstructing found to be easier among the four item types (the literature suggests this item type would be more cognitively demanding than the text-base/literal recall items).

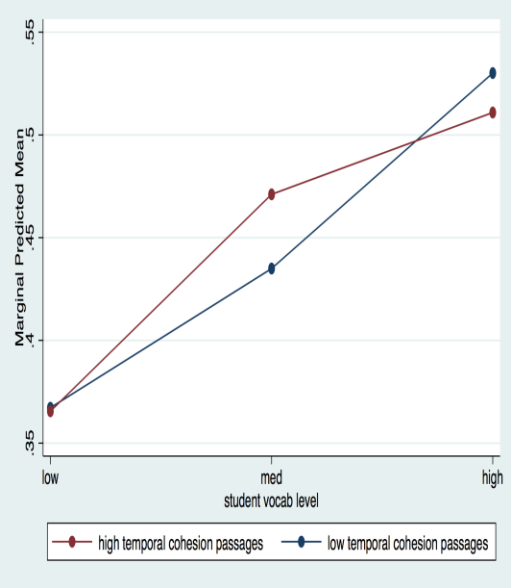
The final set of analyses examined the reader-text and the text-item interactions for a set of text-features that were found to be statistically significant. Some of the results are shown in Figure 4. Interestingly, the text's vocabulary demand, as measured by the mean log word frequency, gives a bigger boost for students with higher vocabulary level while the extent to which the text includes temporal connectives (e.g., first, until, finally) helps students with the average vocabulary level more.

	Est.	SE
Fixed Effects		
Text Representation (TR)		
Ave. sent length	**0.24*** **	**0.04**
Log mean word freq	**-0.11*** **	**0.05**
Narrativity	-0.06	0.06
Syntactic simplicity	-0.16	0.06
Word concreteness	>0.01	0.03
Referential cohesion	-0.03	0.03
Deep cohesion	-0.01	0.02
Verb cohesion	-0.02	0.02
Connectivity	0.04*	0.02
Temporality	**-0.07****	**0.02**
Response Decision (RD)		
Vocab Demand, Correct Ans.	0.03	0.01
Item Type (ref = text-base)		
reword/restructure	**-0.11****	**0.05**
bridging	-0.08	0.06
inferential	0.05	0.05
Abstractness of Info (ref=highly concrete)		
somewhat concrete	**0.21*** **	**0.05**
somewhat abstract	**0.13*** **	**0.04**
highly abstract	**0.09****	**0.05**
Reader		
Vocabulary level	**0.33*** **	**0.02**
Intercept	**-0.20****	**0.07**
Random Effects		
reader variance $\sigma^2_{\epsilon_\theta}$	**0.44****	**0.02**

F*igure 3*. Parameter estimates from the text representation (TR) and the response decision (RD) combined model. Bolded are the results also replicated in the cross- validation sample.



A. LWF* x student vocabulary



B. Temporal cohesion x student vocabulary

Figure 4. The text-reader interactions: Panel A shows the interaction between the mean log word frequency (MLWF) and student vocabulary, while Panel B shows the one between the temporality and student vocabulary.

Conclusion

This study illustrates how explanatory item response models can be used to directly model the three factors, namely the reader, the text, and the items/tasks that affect reading comprehension in the multiple-choice testing context. When interaction terms are included, the models help untangle interesting interplays among the passage, items, and the reader, facilitating a better grasp of both inter- and intra-individual differences (Snow, 2002). The findings resulting from these models have great potential in contributing to better measurement of, and instruction for, reading comprehension across different stages of development.

Yukie Toyoma
University of California, Berkeley

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response theory models: A generalized linear and nonlinear approach*. New York, NY: Springer.

Embretson, S. E. (1983). *An incremental fit index for the linear logistic latent trait model*. Paper presented at the annual meeting of the Psychometric Society, Los Angeles.

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. *Applied*, *11*(2), 175–193.

Gorin, J. S., & Embretson, S. E. (2006). Item difficulty modeling of paragraph comprehension items. *Applied Psychological Measurement*, *30*(5), 394–411.

Kincaid, J. P., Fishburne, L. R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. In *Research Branch Report* (pp. 8–75). Memphis: Chief of Naval Technical Training: Naval Air Station.

Mislevy, R. (1987). Exploiting Auxiliary Infornlation About Examinees in the Estimation of Item Parameters. *Applied Psychological Measurement, 11*(1), 81–91.

Mislevy, R., & Sheehan, K. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika, 54*(4), 661–679.

National Governors Association Center for Best Practices (NGACBP) & Council of Chief State School Officers (CCSSO). (2010). *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects with Appendices A–C*. Washington DC.

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: the passage or the question? *Behavior Research Methods, 40*(4), 1001–1015.

Sheehan, K. M., Kostin, I., Napolitano, D., & Flor, M. (2014). The TextEvaluator tool : Helping teachers and test developers select texts for use in instruction and assessment. *Elementary School Journal*, 115(2), 184–209.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6(2), 461-464.*

Journal of Applied Measurement
Vol. 19, No. 3, Fall 2018

The Impact of Levels of Discrimination on Vertical Equating in the Rasch Model - *Stephen M. Humphry*

Person-Level Analysis of the Effect of Cognitive Loading by Question Difficulty and Question Time Intensity on Didactic Examination Fluency (Speed-Accuracy Tradeoff) - *James J. Thompson*

Detecting Rater Effects under Rating Designs with Varying Levels of Missingness - *Rose E. Stafford, Edward W. Wolfe, Jodi M. Casabianca, and Tian Song*

A Rasch Model Analysis of the Emotion Regulation Questionnaire - *Michael J. Ireland, Hong Eng Goh, and Ida Marais*

Measuring the Impact of Caring for a Spouse with Alzheimer's Disease: Validation of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQue) - *Peter Hagell, Matthew Rouse, and Stephen P. McKenna*

Psychometric Properties of an Instrument to Measure Word Problem Solving Skills in Mathematics - *Chunlian Jiang, Do-Hong Kim, and Chuang Wang*

Evaluation of Self-Management Program Outcomes: Adaptation and Testing of a Swedish Version of the Health Education Impact Questionnaire (heiQ) - *Christine Kumlien, Michael Miller, Cecilia Fagerström, and Peter Hagell*

Developing and Validating a Scientific Multi-Text Reading Comprehension Assessment: In the Text Case of the Dispute of whether to Continue the Fourth Nuclear Power Plant Construction in Taiwan - *Lin Hsiao-Hui and Yuh-tsuen Tzeng*

Richard Smith, Editor, www.jampress.org

Infographs, Isotopes and Measurement: The Genius of Otto Neurath and Gerd Arntz

In recent years, the use of infographics has emerged as a popular technique for visually communicating data. Infographics may be readily found online on informational websites (e.g., the Centers for Disease Control, National Institutes of Health, etc.) and in health and educational settings, such as hospitals, clinics, and classrooms. However, long before modern infographics with their stunning visuals and text that meticulously adheres to best practices in visual communication, there was the 'Isotope', an acryonm for International System Of Typographic Picture Education. In short, Isotope refers to a method of utilizing standardized and abstracted symbols to present biological, historical, social and technological data in pictorial form (Bresnahan, 2011).

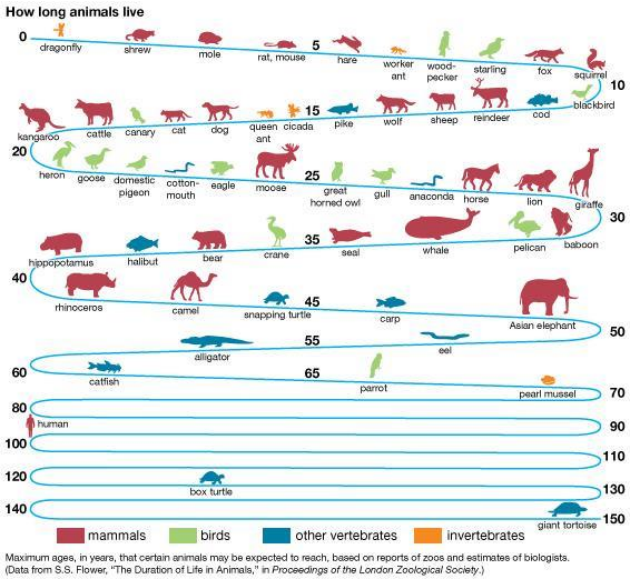


Figure 1. Neurath and Arntz's pictogram.

Otto Neurath, an Austrian philosopher and scientist, was credited with the first isotope when he founded a museum in Vienna, Austria in the 1920s. Unlike most museums that simply hosted valuable objects and relics, Neurath envisioned a teaching museum in which facts and statistics could be presented with pictures, thus making them more attractive and memorable for attendees. Neurath was quoted as stating "To remember simplified pictures is better than to forget accurate figures" (Haller & Kinross, 1991). In the mid-1920s, Neurath sought the assistance of renowned German artist Gerd Arntz to bring ideas to life with visuals. Arntz would go on to be credited with more than 4,000 symbols and figures (Annick & Bruinsma, 2010).

One interesting example of Neurath and Artnz's work is the "How Long Animals Live" infographic (see Figure 1). Although there are certain measurement properties absent from this visual (e.g., scaling of animal size, interval level scaling of the y-axis, etc.), it nonetheless serves as an early, rudimentary example of an infographic that effectively displays its intended message while mostly retaining the properties of objective measurement. This infographic is particularly noteworthy for modern measurement scholars looking for creative ways to present data.

Kenneth Royal
North Carolina State University

References

Bresnahan, Keith (2011). "An Unused Esperanto": Internationalism and Pictographic Design, 1930–70". Design and Culture, 3(1).

Otto Neurath, Gesammelte Bildpädagogische Schriften [ed. Rudolf Haller & Robin Kinross]. Vienna, Hölder-Pichler-Tempsky, 1991. pp. 5, 57, 132.

Annink, Ed; Bruinsma, Max (2010). Gerd Arntz Graphic Designer. Rotterdam: 010 Publisher

An Alternative Interpretation of the Recent Pew Poll Contrasting Differences as to the "Very Big" Problems We Face Today

An online news item appearing on 15 October 2018 proclaims that "Americans don't just disagree on the issues. They disagree on what the issues are." The article, by Dylan Scott on the Vox website, reports on a poll conducted by the Pew Research Center, involving registered voters in the U.S., between 24 September and 7 October. Polarizing disagreement is a recurring theme in the world, and keeping the tension up sells ads, so it is not surprising to see the emphasis in both the article and in the Pew report on major differences in people's perceptions of what counts as a "very big" problem in the U.S. today. But a closer look at the data gives hope for finding ways to communicate across barriers that may look more significant than they actually are.

There's no mention in the article of the sampling error, uncertainty, or confidence level, but the Pew site indicates that, overall, sampling error is 1.5%. But the Vox article mentions only the total sample size and fails to say that the registered voter portion of the respondents is smaller by a couple of thousand. Further, the sampling error jumps up to 2.6% for respondents indicating support for a Republican candidate, and to 2.3% for respondents supporting a Democrat. Again, the differences being played up are quite large, so there's little risk of making too much out of a small difference. It's good to know just how much of a difference makes a difference, though. That said, neither Pew nor the Vox story mentions the very strong agreement between the different groups supporting opposing party candidates when the focus is on the relative magnitudes of agreement on aligned issues. Survey research typically focuses, of course, on percentages of responses to individual questions. Only measurement geeks like me wonder whether questions addressing a common theme could be related in a way that might convey more information. My curiosity was piqued, even though it is impossible to properly evaluate a model of this kind from the mere summary percentages. I knew if I found any correspondences they might just be accidents or coincidences, but I wanted to see what would happen.

So I typed up the text of the 18 issues concerning the seriousness of the problems being confronted in the US today, along with the percentages of registered voters saying each is a "very big" problem today. I put it all into SPSS and made a few technical checks to see if any major problems of interpretation would emerge from the nonlinear and ordinal percentages. The plots and correlations I did indicated that the same general results could be inferred from both the Pew percentages and their logit transformations. While I was looking at a scatter plot of the Republican vs Democrat agreement percentages I noticed something interesting. I had been wondering if perhaps the striking differences in the groups' willingness to say problems were serious might be a matter of relative emphases. Might the Republican supporters be less willing to find anything a big problem, but to nonetheless rank the issues in the same order as the Democrat supporters? This is, after all, exactly the kind of pattern commonly found in data from various surveys, assessments, and tests. No matter whether a respondent scores low overall, or scores high, the relative order of things stays the same.

Now, this is true in the kind of data I work with because considerable care is invested in composing questions that are intended to hang together like that. The idea is to deliberately vary the agreeability or difficulty of the questions so they all tap the same basic construct and demonstrably measure the same thing. When these kind of data are obtained, different questions measuring the same thing can be asked of different people without compromising the unit of measurement. That is, each different examinee or respondent can answer a unique set of questions and still have a measure comparable with anyone else's. Like I said, this does not just happen by itself, but has to come about through a careful process of design and calibration.

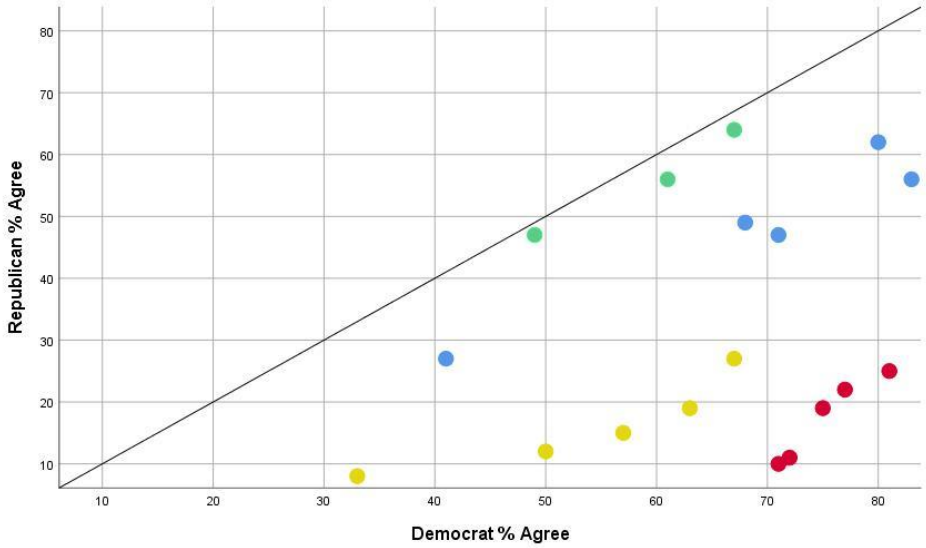


Figure 1 Initial plot of Republican vs Democrat Percentages agreement as to "Very Big" problems

But the basic principles are well-established as being of longstanding and proven value across wide areas of research and practice.
So I was wondering if there might be one or more subsets of questions in the Pew data that would define the same problem magnitude dimension for supporters of both Republican and Democratic candidates. And as soon as I looked at the scatterplot of the percentages from the two groups, I saw that yes, indeed, there appeared to be four groups of issues that lined up along shared slopes. A color-coded version of that plot is in Figure 1.

The one statistical inference problem that emerged in examining these ordinal data concerns the yellow dot that is lowest and furthest to the left. At 8% agreement from the Republican supporters it was pulled away from the linear relation further than the other correspondences. When transformed into a log-odds unit, that single problematic difference lines up well with the other yellow dots further to the right.

The identity line in the figure shows where exact agreement between the two groups would be. That line marks out the connection between the same percentages of respondents agreeing an issue is a "very big" problem. We can see that the three green dots very nearly fall on that identity line. Just below them is a row of blue dots almost parallel with the identity line. Then there's a third row of yellow dots further down, indicating more absolute disagreement between the two groups on these issues, but also showing a quite strong agreement as to their relative magnitudes within that group. Finally, there is another, red, line of dots in the lower right corner of the figure that marks out a more extreme range of absolute disagreement, but is also quite parallel to the identity line.

Figures 2-5 illustrate each of these groups of issues separately, giving further information on the problems and showing the regression lines and correlations for each contrast. The same colors have been retained to aid in seeing which groups of issues in Figure 1 are being shown. The four areas of problems seem to me to correspond to issues of perceived major threats (Figure 2), accountability and access issues (Figure 3), equal opportunity issues (Figure 4), and systemic problems (Figure 5). Each of these content areas could be explored conceptually and qualitatively to assess whether some initial sense of a measured construct can be formed. If the by-person individual response data could be analyzed for fit to a proper measurement model, a much better job of determining the presence of invariant structure could be done.

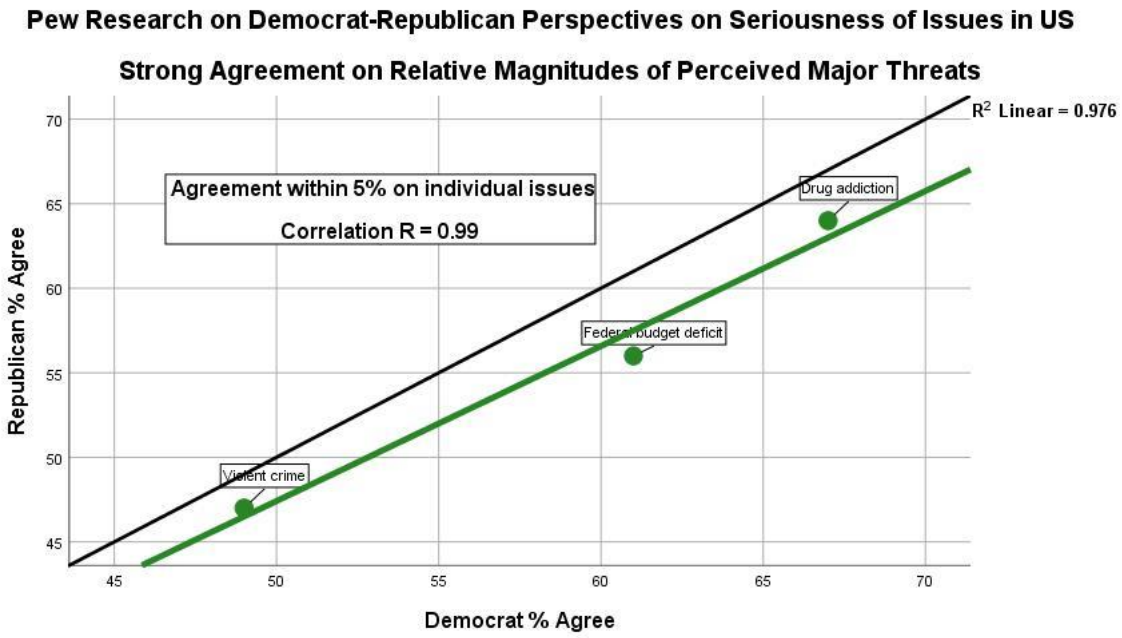


Figure 2.Republican vs Democrat areas of agreement as to "Very Big" problems

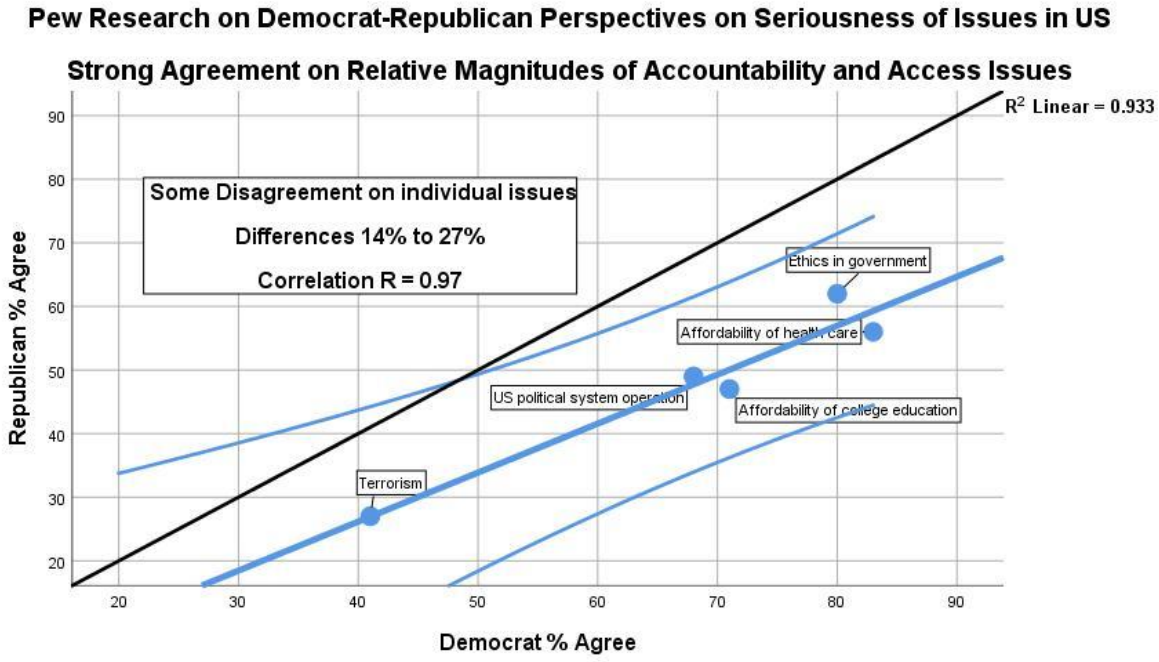


Figure 3 Republican vs Democrat areas of some disagreement as to "Very Big" problems

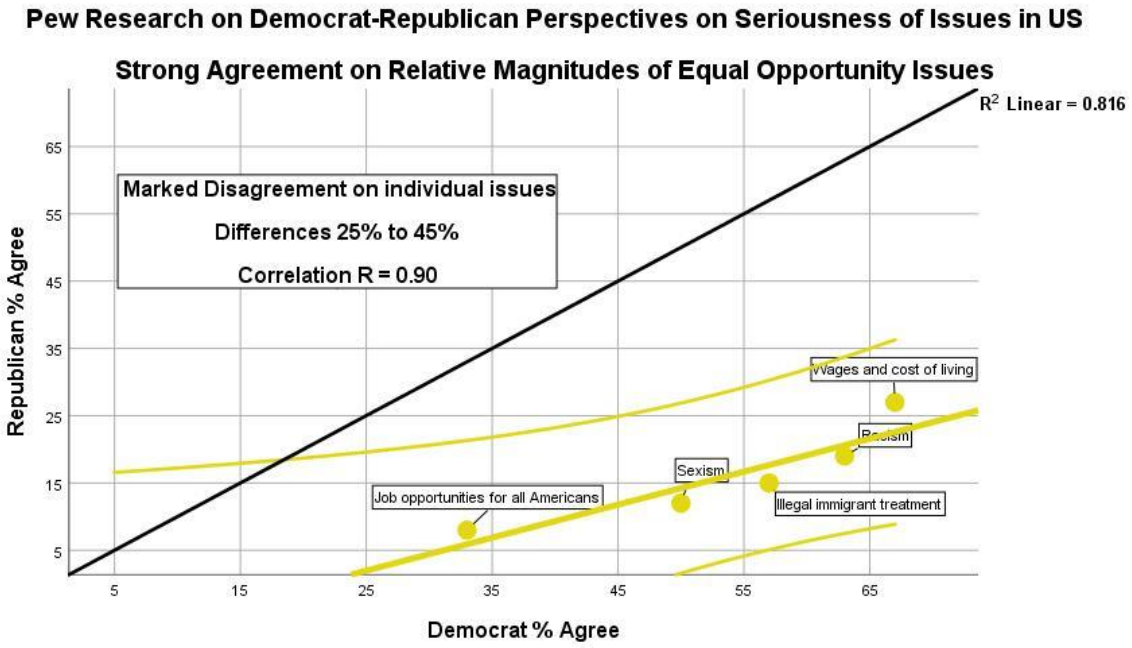


Figure 4 Republican vs Democrat areas of marked disagreement as to "Very Big" problems

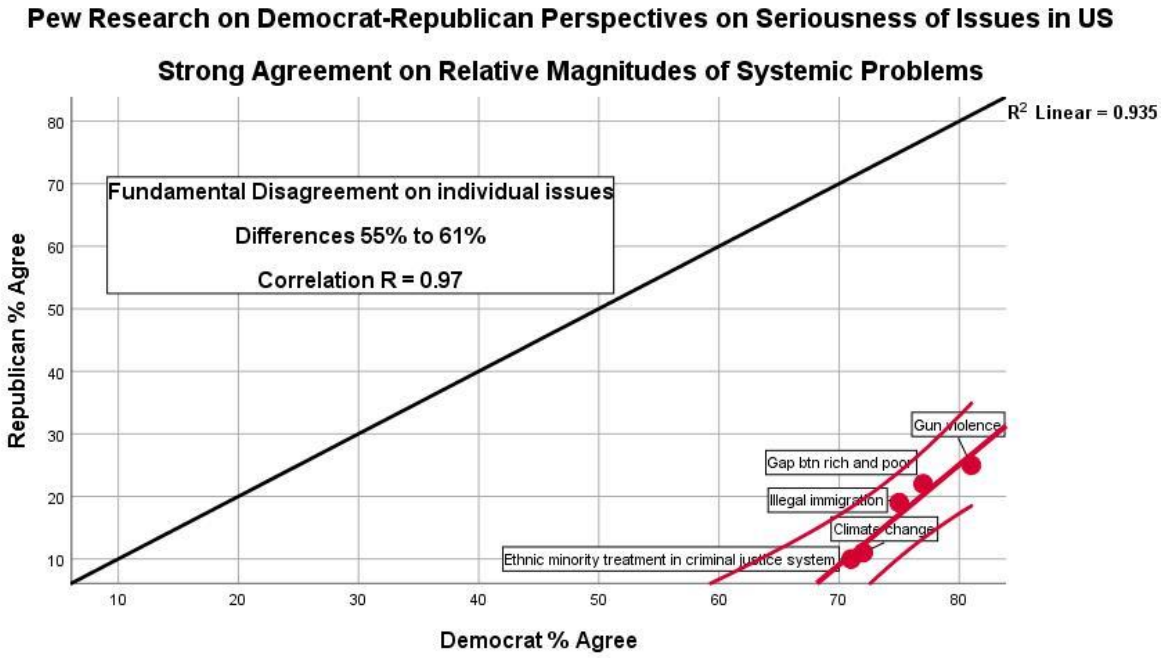


Figure 5 Republican vs Democrat areas of fundamental disagreement

But even without undertaking that work, these results already suggest a basis for productive conversations between the supposedly polarized groups. To start from the low-hanging fruit, the three problems the two groups agree on to within a couple of sampling errors (Figure 2) present topics of common agreement. Both Democrats and Republicans identify violent crime, the

federal budget deficit, and drug addiction as mattters of equally shared concern. The point is not that these are the highest rated problems for either group, but, rather, that they agree within the limits of statistical precision as to the extent that these are "very big" problems. It may be that setting shared priorities for addressing these problems could ground new relationships in that experience of having accomplished something productive together.

This new approach to building social capital might then proceed by taking up progressively more difficult areas of disagreement as to what "very big" problems are. Even though Republicans rate each area as less likely to be a "very big" problem, within each of the four groups of issues, they agree with Democrats as to their relative magnitudes. News like this might not sell a lot of ads, but it does offer hope for finding new ways of approaching relationships and crossing divides.

William P. Fisher, Jr.
Sausalito, CA

Ten Axes of the Rasch Model

Ten axes of Rasch model shown in Figure 1 was presented by Professor Wen-Chung Wang at PROMS 2016 in Xi'an, China. Although Professor Wen-Chung passed away on Feb. 15, 2018(Cavanagh, 2018), the online presentation leaves us and guides Rasch learners an overall perspective to know the evolution and extensions of Rasch model. We provide readers with videos (Chien & Shao, 2018) online in honor of Wang's contribution to the development of Rasch model in his lifetime (Chien, 2018).

We listed the ten axes of Rasch model WANG's presented as below:

1. **Response category** includes (1) ordered category, such as Rating scale model(Andrich, 1978) and Partial credit model(Masters, 1982); (2)partial ordered category as ordered partition model(Wilson, 1992);(3) multiple choice items as Rasch model for MC items(Wang, 1998); (4) Continuous response(Mueller,1987).

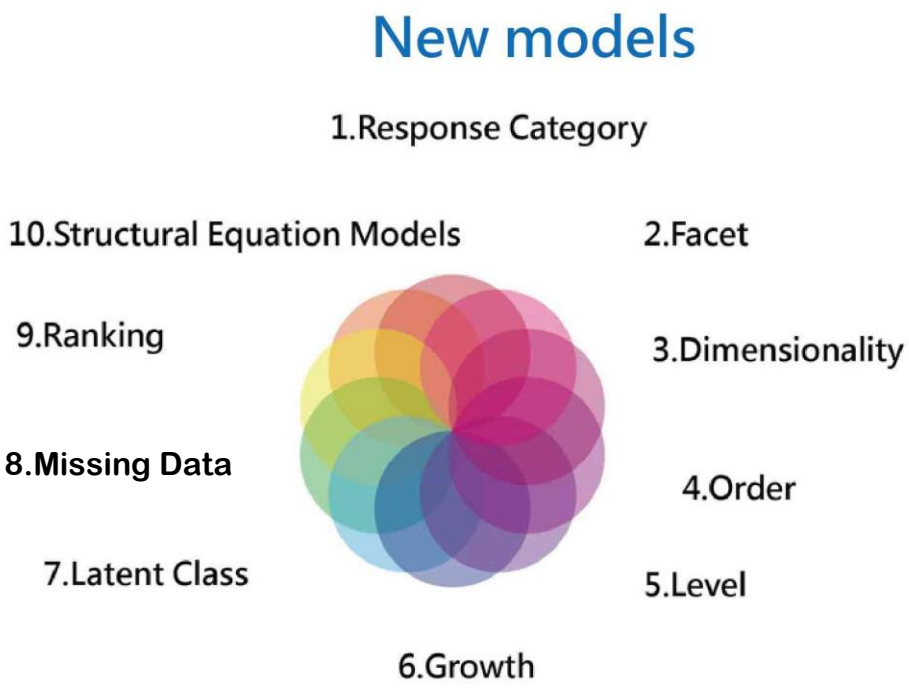
2. **Facet** comprises (1) Linear logistic test model: Item difficulty is decomposed into several components(Fischer, 1973); (2)Many faceted model: Two facets are extended to more than two facets(e.g., rater facet)(Linacre, 1989); (3) Explanatory models: Covariates in item and persons(de Boeck & Wilson, 2004).

3. **Dimensionality** is to measure a very specific attribute with a high degree of accuracy or to sample a vaster range of attributes with much less accuracy. It would be great if test reliability can be increased by using more efficient statistical methods, and correlations between latent traits obtained from unidimensional approaches would be attenuated by measure error(Adams, Wilson, & Wang, 1997). There are three types such as (1)Between-item Multi-dimentionality: A test with several subtests, and each subtest measures a distinct latent trait; (2)Within-item Multidimensionality: A problem-solving task can measure both "content knowledge" and "skill; (3)Testlet(Bi-factor) models (Wang & Wilson,2005).

4. **Order** (Huang, Wang, Chen, & Su, 2013) is used for the example of language proficiency(i.e., second-order latent trait) that often covers scopes of listening, speaking, reading, and writing.

5. **Level** (Huang & Wang, 2014) describes multi-staged sampling scenario such as school are sampled first, followed by student sampling(e.g., PISA).

6. **Growth** consists of **(1)** Individual difference in change(Andersen,1985; Embretson,1991); (2) Multi-level approach is dealing with within occasion, between-occasion, and between persons (Huang & Wang, 2012; Wang, Wilson, & Adams,1998).



7. **Latent Class** explores unknown subpopulations such as Developmental change(Wilsonm 1989), Response styles(Rost et al., 1997), Performance decline during testing(Jin & Wang, 2014), DIF analysis (Frick et al., 2015; von Davier & Carstensen, 2007).

8. **Missing data** handles the scenario of missing not at random, such as survey data (Holman & Glas, 2005), examinee-selected items(e.g., select 2 from 5 given items)(Wang, Jin, Qiu, & Wang, 2012; Wang & Liu, 2011), and "Don't know" option(Liu & Wang, 2016).

9. **The ranking** is attributed to (1) pairwise-comparison items on which one if more like you? e.g., smart or easygoing and lazy or picky were forced to choose anyone from the two; (2) Ranking items are to rank the three statements according to your work interest: for your interest choosing one of them (1) Maintain, install, or repair computers; (2) teach people new skills; (3) Write books, articles, essays, or plays.

10. **Structural Equation Models** are dealing the relationship between indicators(item responses) and latent traits that should be nonlinear. When using tetrachoric or polychoric correction that CANNOT resolve the problem, as did in LISREL. By contrast, latent response approach or IRT is more appropriate for facing this harsh situation (Skrondal & Rabe-Hesketh,2004).

Tsair-Wei Chien
[1]Chi Mei Medical Center, Taiwan

Yang Shao
Tongji Zhejiang College, Jiaxing, China

References

Adams, R. J., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*,1-23.

Andersen Erling B. (1985) Estimating latent correlations between repeated testings. *Psychometrika 50,* 3-16

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

Cavanagh, R.F.(2018). In Memoriam: Professor Wang Wen Chung. *Rasch Measurement Transactions,31,*4,1662.

Chien, T.W.& Shao, y.(2018). Ten axes of Rasch model presented by Wen-Chung Wang. 2018/5/2 retrieved at https://www.youtube.com/watch?v=oe3Tlsc_uhI and https://www.youtube.com/watch?v=zTmlbTfOAnk

Chien, T.W. (2018). WANG's paper publications shown online on Google maps. 2018/5/2 retrieved at http://www.healthup.org.tw/gps/wangpublication.htm

De Boeck and Wilson (2004). *Explanatory item response models: A generalized linear and nonlinear approach.* New York: Springer-Verlag.

Embretson S.E. (1991) A multidimensional latent trait model for measuring learning and change. Psychometrika, 56, 495-515.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37,*359-374.

Frick H, Strobl C, Zeileis A (2014). Rasch mixture models for DIF Detection: A comparison of old and new score specifications. *Educational and Psychological Measurement, 75*(2),208–234.

Holman, R.,& Glas, C.A.W. (2005). Modeling nonignorable missing data mechanism with item response theory models. *British Journal of Mathematical and Statistical Psychology, 58,* 1-18.

Huang, H-Y., Wang, W.-C., Chen, P.-H., & Su, C.-M. (2013). Higher-order item response models for hierarchical latent traits. *Applied Psychological Measurement, 37,* 619-637.

Huang, H.-Y., & Wang, W.-C. (2014). Multilevel higher-order item response theory models. *Educational and Psychological Measurement, 74,* 495-515.

Jin, K.-Y., & Wang, W.-C. (2014). Item response theory models for performance decline during testing. *Journal of Educational Measurement, 51,* 178-200.

Liu, C.-W., & Wang, W.-C. (2016). Unfolding IRT models for Likert-type items with a don't know option. *Applied Psychological Measurement, 40,* 517-533.

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

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

Mueller, H. (1987). A Rasch model for continuous ratings. *Psychometrika, 52,* 165-181.

Rost, J., Carstensen, C., & von Davier, M. (1997). Applying the mixed Rasch model to personality questionnaires. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 324-332). Kiel, Germany: Leibniz Institute for Science and Mathematics Education. Retrieved from http://www.ipn.unikiel.de/aktuell/buecher/rostbuch/ltlc.htm

Skrondal A. and Rabe-Hesketh S. (2004) Generalized latent variable modeling: multilevel, longitudinal and structural equation models. *Statistical Methods in Medical Research, 17* (1). p. 119.

Wang, W.-C., Wilson, M., & Adams, R. J. (1998). Measuring individual differences in change with Rasch models. *Journal of Outcome Measurement, 2,* 240-265.

Wang, W.-C. (1998). Rasch analysis of distractors in multiple-choice items. *Journal of Outcome Measurement, 2,* 43-65.

Wang, W.-C., & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement, 29,* 126-149.

Wang, W.-C., & Wilson, M. (2005). Exploring local item dependence using a random-effects facet model. *Applied Psychological Measurement, 29,* 296-318.

Wang, W.-C., Liu, C.W. (2011).The random-threshold generalized unfolding model and its application of computerized adaptive testing. *Educational and Psychological Measurement 71* (1), 114-128

Wang, W.-C., Jin, K.-Y., Qiu, X.-L., & Wang, L. (2012). Item response models for examinee-selected items. *Journal of Educational Measurement, 49,* 419-445.

Wang, W.-C., Su, C.-M., & Qiu, X.-L. (2014). Item response models for local dependence among multiple ratings. *Journal of Educational Measurement, 51*, 260-280.
von Davier & Claus H. Carstensen (Eds.) (2007). *Multivariate and Mixture Distribution Rasch Models*. New York, NY: Springer
Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin,105*, 276–289.
Wilson, M. (1992). The ordered partition model: An extension of the partial credit model. *Applied Psychological Measurement, 16*, 309–325.

Update on Rasch in Metrology

The International Measurement Confederation (IMEKO) World Congress, was held 3-6 September in Belfast, Ireland (http://www.imeko2018.org/), and featured a special session on Rasch measurement. This session was in the program as part of the efforts of Technical Committee 7 (TC7) on Measurement Science; the contributions of Sanowar Khan and Eric Benoit to the formation of the special session are especially acknowledged.

Mark Wilson chaired, and six presentations were be authored and co-authored by (among others) David Andrich, Stefan Cano, Robert Cavanagh, William Fisher, Luca Mari, Andrew Maul, Jeanette Melin, Leslie Pendrill, Thomas Salzberger, Jack Stenner, and Mark Wilson.

A critical perspective on the six papers was be offered by longtime TC7 member, Giovanni Rossi, author of "Measurement and Probability: A Probabilistic Theory of Measurement with Applications" (Springer, 2014; https://www.springer.com/us/book/9789401788243).

A number of other Rasch papers wer be presented at the World Congress under the auspices of other Technical Committees, such as TC1 on Metrology Education, TC13 on Measurements in Health and Medicine, and TC18 on Human Measurements. More information on these will be available on the conference web site when the full program is released.

TC18 on Human Measurements will be of special interest as it has been invited to participate for the first time in the Joint Symposium with TC1, TC7, and TC13 next year in St. Petersburg, Russia. TC18 could become a group of central interest as national metrology institutes around the world begin to investigate the possibilities concerning unit standards in education, health care, and other areas (as is already underway in Europe; see https://www.ri.se/kalendarium/2017-07-03-empir-15hlt04-neuromet-meeting-mini-symposium).

IMEKO members are the national metrology institutes from countries around the world. The focus of these institutes is on various activities related to the maintenance and improvement of the International System of Units (SI), commonly called the metric system. Rasch measurement has been represented at various IMEKO World Congresses and Joint Symposia at least since 2008.

The IMEKO journal, *Measurement*, published by Elsevier, has featured a number of papers authored jointly by metrology engineers and psychometricians, (Mari & Wilson, 2014; Mari, Maul, Torres Irribarra, & Wilson, 2016; Maul, Mari, Torres Irribarra, & Wilson, 2018; Pendrill & Fisher, 2015), as well as by psychometricians independently presenting the case for a metrological understanding of measurement in psychology and the social sciences (Fisher, 2009; Fisher & Stenner, 2016; Maul, Torres Irribarra, & Wilson, 2016; Wilson, 2013).

(Pictured: David Andrich, William Fisher, Jr., and Peter Hagell at the BIPM in Paris)

The proceedings of the IMEKO World Congress and Joint Symposia are published regularly in the *Journal of Physics Conference Series*. At total of over 50 Rasch-oriented papers have appeared in JoPCS since 2010, including about 30 that were presented at the 2016 Joint Symposium hosted by Wilson and Fisher at UC Berkeley (Wilson & Fisher, 2016).

In addition to IMEKO, Rasch measurement was represented in the larger world of metrology at the September 2017 International Congress of Metrology (CIM) held in Paris. Presentations authored by Leslie Pendrill, Jack Stenner, Matt Barney, and William Fisher were the sole instances of psychological or social metrology at this industry convention, attended by technology companies providing precision instruments. As Leslie Pendrill is a past chair of the organizing committee for this CIM event, he is well known to attendees, who focused significant attention on the new opportunities for the advancement of science represented by Rasch measurement theory. For more information, see http://cim2017.com/programme-cim-2017.html.

Finally, Rasch measurement was strongly represented as well at the recent Measurement at the Crossroads conference held in Paris in June, 2018. Papers by Eran Tal, Josh McGrane, Alex Scharaschki, and Trisha Nowland augmented talks authored and co-authored by Andrich, Cano, Fisher, Mari, Maul, Melin, Pendrill, and Wilson. For more information, see https://measurement2018.sciencesconf.org/resource/page/id/5.

Much remains to be done to elaborate the theory and practice of metrology and measurement for psychology and the social sciences, but the long history of significant work in the field is beginning to be recognized for its value.

William P. Fisher, Jr.
BEAR Center, Graduate School of Education, University of California, Berkeley
wfisher@berkeley.edu

References

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

Fisher, W. P., Jr., & Stenner, A. J. (2016). Theory-based metrological traceability in education: A reading measurement network. Measurement, 92, 489-496.

Mari, L., Maul, A., Irribara, D. T., & Wilson, M. (2016). Quantities, quantification, and the necessary and sufficient conditions for measurement. Measurement, 100, 115-121.

Mari, L., & Wilson, M. (2014). An introduction to the Rasch measurement approach for metrologists. Measurement, 51, 315-327.

Maul, A., Mari, L., Torres Irribarra, D., & Wilson, M. (2018). The quality of measurement results in terms of the structural features of the measurement process. Measurement, 116, 611-620.

Maul, A., Torres Irribarra, D., & Wilson, M. (2016). On the philosophical foundations of psychological measurement. Measurement, 79, 311-320.

Pendrill, L., & Fisher, W. P., Jr. (2015). Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number. Measurement, 71, 46-55.

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

Wilson, M., & Fisher, W. (2016). Preface: 2016 IMEKO TC1-TC7-TC13 Joint Symposium: Metrology Across the Sciences: Wishful Thinking? Journal of Physics Conference Series, 772(1), 011001.

Teaching about Rasch

As I teach about Rasch, I am always interested in novel applications of Rasch methods. I have for a long time felt that Rasch would be of immense use in the business world. Below I provide a citation for a thoughtful article that used Rasch methods to evaluate credit ratings. The authors (Gori & Gori, 2018) state in their abstract: "The purpose of this paper is to understand if the Rasch model can be applied to mimic the credit ratings and can help to develop a simple and objective way to evaluate the creditworthiness of companies and their financial obligations".

I am adding this to my class reading list to help my students better see the use of Rasch in many fields.

Bill Boone
Miami University
boonewjd@gmail.com

Reference

Gori, E. & Gori, G. (2018). Credit ratings: A new objective method using the Rasch model: The case of consumer discretionary. *International Journal of Business and Management Studies, 7*(1) 53-84.

Upcoming Conference Proposal Deadlines

The 2019 International Measurement Confederation (IMEKO) TC1-TC7-TC13-TC18 Joint Symposium will be held 2-5 July in St. Petersburg, Russia. The deadline for extended abstract submissions is 15 November (https://imeko19-spb.org/submission-of-papers/). Please be sure to format your submission per the Journal of Physics Conference Series requirements, as specified on the web site. Contact Kseniia Sapozhnikova at k.v.s@vniim.ru for more information and questions about possible deadline extensions.

The 19th International Metrology Congress will be held 24-26 September, 2019, in Paris, France. Paper presentation proposals are due by 15 January. For more information, see http://cim2019.com/call-for-papers.html.