RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 27 No. 1 | Summer 2013 | ISSN 1051-0796

An "Estimation Bias" Shootout in the Wild West: CMLE, JMLE, MMLE, PMLE

Three riverboat gamblers, Bruce, George and Ben, are discussing tomorrow's sharpshooting contest between Annie Oakley and Lillian Smith. Today's *Deadwood Pioneer* newspaper contains reports of their previous contests in Table 1 and of their contests with Frank Butler in Table 2.

Table 1. Previous Contests between Annie and Lillian. (1 = Winner)

Year	Annie	Lillian
1888	1	0
1887	0	1
1886	1	0
Shooter's Score	2	1

Table 2. Contests between Annie or Lillian and Frank. (1 = Winner)

Year	Annie	Lillian
1888	1	0
1887	0	1
1886	1	0
Shooter's Score	2	1

"Tomorrow it's Annie against Lillian. Here's how to get the odds correct." says Bruce. "Let's use PMLE[1]. In each Table, Annie and Lillian were in the same situation three times, once for each row. Annie won twice. Lillian won once. The odds in both Tables are 2/1. Annie and Lillian are ln(2/1) = 0.69 logits apart."

"We get those same odds of 2/1 from both Tables using CMLE[2,3]." agrees George.

"JMLE[4] and MMLE[5] estimate that the odds for both Tables are 4/1" says Ben. "To produce the correct odds of 2/1 for the direct pairwise comparison of Annie and Lillian in Table 1, we must adjust for JMLE estimation bias[6]. However, the odds for the indirect pairwise comparison of Annie and Lillian in Table 2 are 4/1. JMLE/MMLE are correct. There is no JMLE estimation bias for Table 2."

"No! No!" objects George. "JMLE estimates are always biased[7], even though the bias reduces quickly for larger datasets[8]. Ben, you are way off target!"

Table 3. Contests between Annie or Lillian and Frank (redrawn). (1 = Winner)

Year	Annie	Frank	Lillian
1888	1	0	
1887	0	1	
1886	1	0	
1888		1	0
1887		0	1
1886		1	0
Shooter's Score	2 of 3	3 of 6	1 of 3

Table of Contents

"It's you guys who can't shoot straight" says Ben. "Let's redraw Table 2 so that each row is a separate contest of sharpshooters. Here it is in Table 3 where all the participants are columns and there is one row for each pairwise contest, exactly like Table 1. Let's take Bruce's PMLE logic for Table 1 and apply it to Table 3. All the row scores are 1. In the upper three contests, Annie scores 2 and Frank scores 1. The odds are 2/1 for Annie against Frank. In the lower three contests, Frank scores 2 and Lillian scores 1. The odds are 2/1 for Frank against Lillian. Combining these, the odds for Annie against Lillian are (2/1) * (2/1) = 4/1. They are ln(4/1) = 1.39 logits apart, exactly as JMLE tell us!"

"That is exciting!" exclaims Bruce. "We can extend Table 3 to much larger competitive situations such as Basketball[9] and Tennis using PMLE or bias-adjusted JMLE[10]."

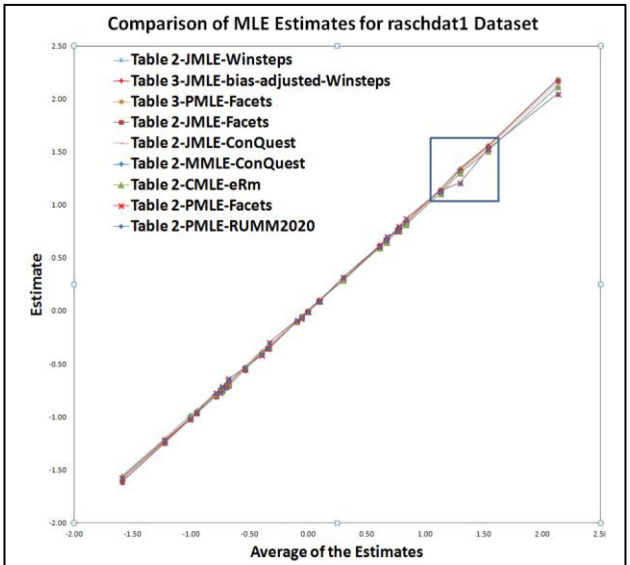


Figure 1. MLE estimates for raschdat1

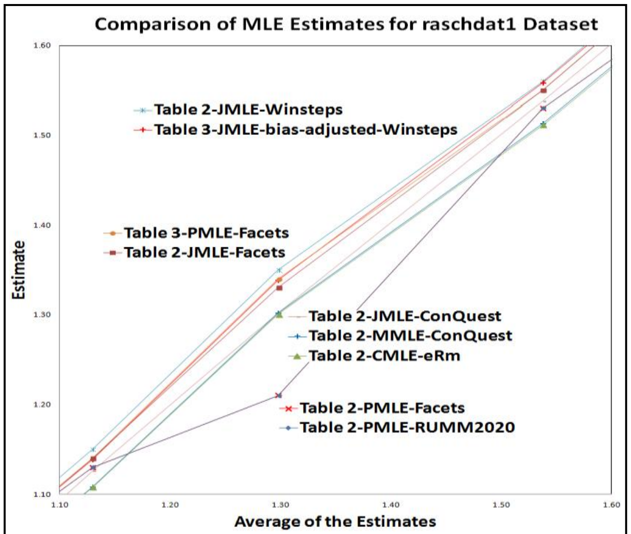


Figure 2. MLE estimates for raschdat1 (Excerpt).

"Ben, why didn't you explain this to me years ago?" says George. "In Tables 1 and 3, the paired comparisons are direct. In Table 2 the comparisons are indirect. Table 2 is an abbreviated version of Table 3. When we treat the comparisons in Table 2 as direct, we distort the meaning of the data, resulting in biased estimates."

"Exactly!" says Ben, "When a dataset is directly pairwise, as in Tables 1 and 3, CMLE/PMLE estimates are accurate and unbiased. We must bias-adjust JMLE estimates. For a dataset of indirect comparisons like Table 2, JMLE estimates are unbiased. CMLE/PMLE estimates for any dataset are biased if reformatting that dataset to be directly pairwise produces different CMLE/PMLE estimates."

Dear Reader: Would you like more evidence? raschdat1.rda is a dichotomous dataset of 30 items and 100 persons distributed in the *eRm* package. In www.rasch.org/rmt/a/shootout.zip, there are conventional Table 2 (30x100) versions of raschdat1 for CMLE, JMLE, MMLE, and PMLE, also Table 3 pairwise (130x3000) versions for JMLE and PMLE, together with their estimates (see Figure 1) and the Excel worksheet of the Figures. In Figure 2, Table 3 curves track with Table 2 JMLE curves, confirming Ben's claim that JMLE is unbiased for conventional datasets.

John Michael Linacre
Gregory Chan (RUMM2020 Analyses)
Raymond J. Adams (ConQuest analyses)

References

1 PMLE, Pairwise Maximum Likelihood Estimation, in *RUMM2030* and *Facets.*
2 CMLE, Conditional Maximum Likelihood Estimation, in *eRm.*
3 *CMLE for Tables 1 and 2:* All the rows are scored 1 on the 2 items, so we only need the probabilities for a row score of 1. Let's call P(10) the probability that Annie wins and Lillian loses, then P(01) is the opposite. For each row the total probability for a score of 1 is P(10) + P(01). In each Table, Annie scored 2 in 3 attempts, so 2 = 3 * P(10) / (P(10) + P(01)). Lillian scored 1 in 3 attempts, so 1 = 3 * P(01) / (P(10) + P(01)). Now, divide those two equations, then the odds are P(10)/P(01) = 2/1. ln(2/1) = 0.69 logits.
4 JMLE, Joint Maximum Likelihood Estimation, in *ConQuest*, *Facets* and *Winsteps.* The JMLE estimates are the ones for which the observed marginal score equals the expected marginal score for each row and column.
5 MMLE, Marginal Maximum Likelihood Estimation, in *ConQuest.* MMLE estimates are the ones for which the observed marginal score for each column equals the expected marginal score, and the row parameters are modeled to have a normal distribution.
6 Using *Winsteps, JMLE* pairwise estimation bias is adjusted by *Paired=Yes*

[7] Andersen E.B. (1970) Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society B* 32, 283–301

[8] Wright, B.D. (1988) The efficacy of unconditional maximum likelihood bias correction: Comment on Jansen, Van den Wollenberg, and Wierda. *Applied Psychological Measurement*, *12*, 315-318.

[9] Linacre J.M. (2001) Paired comparisons for measuring team performance. *RMT, 15*:1, 812. www.rasch.org/rmt/rmt151w.htm

[10] Linacre J.M. (1997) Paired comparisons with standard Rasch software. *RMT, 11*:3, 584-5. www.rasch.org/rmt/rmt113o.htm

Interesting Presentation of Rasch Output!

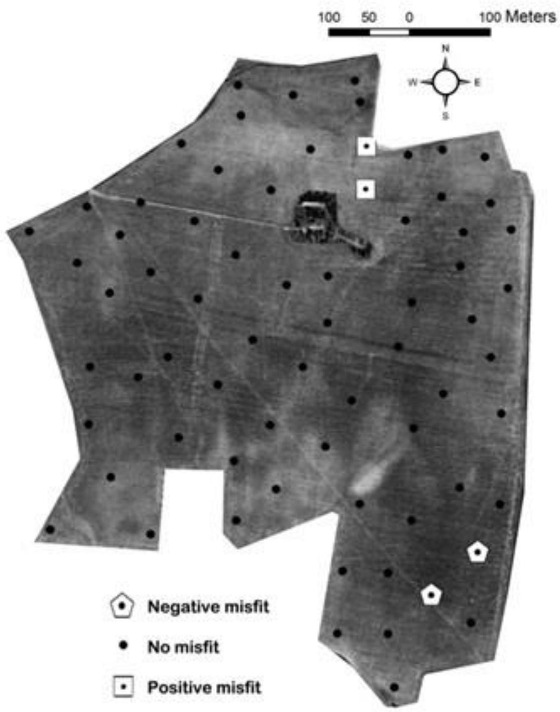


Figure 4: Misfits for clay content in the experimental field.

"In Fig. 4, locations where soil-clay content misfits exist are shown; the two positive and negative misfits are both located together, denoting there is an excess of this textual property in one zone of the field and shortage of the same property in the other zone, with respect to the optimum level to reach a higher soil fertility potential. If it is necessary, any work to amend this soil property should be conducted in these zones". (p. 918).

Moral, F. J., Rebollo, F. J., & Terron, J. M. Analysis of soil fertility and its anomalies using an objective model. *J. Plant Nutr. Soil Sci*. (2012), 175, 912-919. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

Early Detection of Item Miskey on a CAT: The Use of Multiple Indices

An item on an operational test that has been keyed incorrectly represents a threat to score validity. A miskeyed item or items can cause more able examinees to have lower overall scores and less able examinees to have higher overall scores, thus reducing the ability to clearly discriminate between examinees. This is particularly true when test scores are used for classification such as determining whether or not an examinee should be awarded a professional license or a certificate. A procedure that can detect miskeyed items early in an examination cycle improves the integrity of a testing program by reducing the likelihood of misclassifying examinees.

No one statistical index is completely reliable in the detection of a miskeyed item. In classical test statistics, the p-value and the point biserial correlation have frequently been used to identify miskeyed items. A low p-value and a negative point biserial are often interpreted as indicating an item miskey. While these outcomes can indicate an item miskey they are also associated with other item characteristics, such as item ambiguity or multiple correct answers. For a fixed form test, these statistics may be sufficient, however in a computerized adaptive test (CAT) environment the usefulness of the p-value and the point biserial is greatly diminished. CAT examinations are designed to present items to an examinee based on an estimate of the examinee's ability which causes the sample used to calculate p-value and point biserial estimates of items in an operational examination to be different than the reference group used to establish the original item parameters. Further, calculating p-value and point biserial estimates based on a sample obtained from responses generated by a CAT examination, results in statistics which are less stable due to the range restriction of the sample. In item response theory (IRT) models, fit statistics are often used to identify problem items. Commonly the weighted (infit) and unweighted (outfit) standardized mean squares statistics are used to identify items that do not meet the expectations of the measurement model. However, the calculation of both infit and outfit is dependent on deviations from the model expectations and a restricted sample range will also impact these calculations, making it difficult to use them for identifying miskeyed items.

There are three things to consider when identifying miskeys in an operational examination.

1. Is there a statistic or combination of statistics that can identify miskeys?
2. How large of a sample size is needed to create useful decisions?

3. What is the rate of false positive and false negative identifications?

Ideally a single statistic would provide all the information needed to determine a miskey however it may be that a combination of statistics would be more useful. Sample size is important because a method that works well with a smaller sample would enable earlier analysis during a testing cycle reducing the amount of time a miskeyed item was used. Finally, it is important to understand the false positive and false negative rate since too many false positives require manual inspection and too many false negatives would defeat the purpose of the process. Cut-off values can be established for each statistic, such that any item falling above or below an established set of values would be considered to be a likely candidate as a miskeyed item.

To explore this idea a simulation was created to review the performance of readily available statistics to determine if singly or in combination they could provide a consistent identification of miskeyed items. The statistics investigated are p-value, point-measure correlation, infit, outfit, displacement, and upper asymptote. The upper asymptote statistic is available in the Winsteps item analysis and represents a four-parameter IRT model (4-PL) estimate of carelessness or inadvertent selection of a wrong answer. The expectation is that this value should be close to 1 for normal items and much smaller for miskeyed items.

A simulator program (Becker 2012) was used to administer ten replications of a variable length CAT examination, each replication having 1200 examinees. Eight items out of a large pool of items of over 1400 items were selected to be miskeyed items. The simulator program generated test results using the candidate ability measure and the item difficulty to generate a probability of a correct response for each candidate/item interaction. A random number was then generated and, if that number was less than or equal to the probability, the candidate was scored as having answered the question correctly. However, when a candidate encountered a miskeyed item, if the random number was less than or equal to the probability, the candidate was scored as having answered the question incorrectly. The resulting matrix of answer strings was then analyzed using Winsteps and the statistical indices described above were examined to assess their utility in identifying the miskeyed items.

The analysis identified three statistics that, used in combination, gave the cleanest separation between miskeyed and normal items. These statistics were p-value, displacement, and upper asymptote. The cut-off values that were found to be most useful were as follows; p-value <= 0.20, displacement >= 1.5 and upper asymptote <=0.4. Items could receive different N counts based on the selection algorithm used in the variable length CAT. A further cut off was established setting an exposure minimum of 20. The ten replications with eight miskeyed items in each replication presented 80 cases in which a miskeyed item would hopefully be flagged. Using these criteria miskeyed items were flagged in 68 out of the 80 instances (85%). Conversely none of the normal items were flagged out of the 14,640 cases. In the 12 instances in which a miskeyed item was not flagged, 7 involved the same item, which was the hardest item in the miskey set. Logically, hard items are going to be the most difficult to detect as miskeys.

Reference

Becker, K. (2012). Pearson CAT Simulator. Chicago, IL: Pearson VUE.

John A. Stahl and Gregory M. Applegate
Pearson VUE

Investigating Guessing Strategies and Their Success Rates on Items of Varying Difficulty Levels

Psychometricians have long known that guessing is a major threat to the validity of a test score and can be a source for construct irrelevant variance. Guessing behaviors typically are investigated in a number of ways, but almost all involve administering an exam to an appropriate sample and investigating the scores and response patterns for clues that guessing might have occurred. At the University of North Carolina at Chapel Hill, we wanted to evaluate the psychometric integrity of our medical school exam items. In doing so, we opted to construct an exam consisting of actual medical school items and administer them to university staff in the Office of Medical Education. It was theorized that the sample would need to rely almost entirely on guessing strategies as none of the participants had any formal educational or experiential training in medicine or the health sciences. By intentionally offering an exam to an inappropriate sample we were able to more deliberately investigate guessing, identify which exam items were vulnerable to testwiseness, and better discern how guessing might impact the quality of our medical students' test scores.

Study Design

As part of our experiment, a purposeful mix of easy, moderate, and difficult items were randomly pulled from each of the courses that comprise the first two years (pre-clinical) of the medical school curriculum. Criteria for determining easy, moderate, and difficult items were arbitrarily categorized by the following schema. Easy items were those that were answered correctly by 76% or more of medical students; moderately difficult items were those that were answered correctly by 51%-75% of medical students; and difficult items were those that were answered correctly by less than 50% of medical students. The exam consisted of a total of 63 items and was administered to 14 professional staff personnel in the Office of Medical Education. A requirement for participation in the study was that all staff must hold at least a bachelor's degree and have no formal educational or experiential training in the physical, life, or health sciences that might unduly offer an advantage on the exam. These criteria for inclusion were necessary so as to assess primarily guessing behaviors with minimal influence of content knowledge.

Accompanying each item was a follow-up question that asked test-takers to rate the extent to which they relied on guessing strategies to answer the previous question. Using Rogers (1999) framework for guessing, we asked test-takers to indicate whether they relied on random, cued, or informed guessing, or no guessing at all. Specifically, we provided the following item:

Please identify the strategy you used to answer the previous question from the options below:

1) I did not guess.
2) Informed guessing: I selected a particular answer based upon previous partial knowledge of the subject, or I was able to eliminate particular answer options based upon previous partial knowledge of the subject.
3) Cued guessing: I selected an answer based upon some sort of stimulus within the test such as wording cues, cues associated with item stems, choices among answer options, testwiseness, etc.
4) Random guessing: I selected a particular answer by blindly choosing an answer.

Results

Overall, results reveal a mix of guessing strategies were used. Table 1 presents information regarding the use and success of each guessing strategy. Participants reported they did not guess on 17 items, but the success rate for this strategy indicates they were correct only 70% of the time. Random guessing was used most frequently (nearly half the time), but resulted in the lowest success rate (around 24%). Cued and informed guessing resulted in nearly equal success rates (45-49%).

Table 1. *Descriptive Summary of Guessing Strategies and Their Success Rates*

	Type of Guessing Strategy	Guessing Strategy Used		Guessing Correctly		Success Rate
		n	%	*n*	%	%
Overall	No Guessing	17	1.9	12	1.4	70.59
	Informed Guessing	127	14.5	63	7.2	49.61
	Cued Guessing	310	34.9	141	16.1	45.48
	Random Guessing	424	48.2	102	11.6	24.06
Easy	No Guessing	6	2.0	6	4.0	100.00
	Informed Guessing	43	14.6	27	18.0	62.79
	Cued Guessing	110	37.4	76	50.7	69.09
	Random Guessing	133	45.2	41	27.3	30.83
Moderate	No Guessing	8	2.5	5	4.9	62.50
	Informed Guessing	44	13.8	21	20.4	47.73
	Cued Guessing	118	36.9	45	43.7	38.14
	Random Guessing	150	46.9	32	31.1	21.33
Difficult	No Guessing	3	1.0	1	1.5	33.33
	Informed Guessing	40	13.6	15	23.1	37.50
	Cued Guessing	82	27.9	20	30.8	24.39
	Random Guessing	141	48.0	29	44.6	20.57

*Note: There were 4 (.45%) instance of missing data.

To take the analysis a step farther, we investigated guessers' performance based on item difficulty. Using the aforementioned criteria for easy, moderate, and difficult items, guessing strategies were investigated to determine which type of guessing resulted in the best success rate relative to item difficulty. Results indicate the easy items are highly vulnerable to guessing. Such high levels of contamination certainly threaten the validity of the information obtained from these items. Interestingly, cued guessing strategies resulted in a slightly higher success rate on easy items than having informed knowledge. However, as the difficulty of the items increased, success rates between cued and informed guessing strategies tended to shift towards informed guessing providing the greater probability of success. The gap between the success rates of informed guessing over cued guessing also widened when the items became more difficult.

According to Rasch measurement theory, a more knowledgeable person should always have a greater probability of success on any item than someone that is less knowledgeable. Because cued guessing (less knowledge) can result in a greater probability of success on easier items than informed guessing (some partial knowledge), this violates Rasch theory. Results presented here illustrate the necessity for good, sound items that are not susceptible to testwiseness strategies.

Additional Considerations and Recommendations

Guessing can impact virtually any test score. Even the best psychometrically functioning exams result in test-takers having a minimum of 20-25% chance of getting any given item correct when presented with four to five response options. Despite the ever-present threat to validity, it remains unclear to what extent guessing threatens the validity of test scores for persons/organizations that do not have a great deal of psychometric expertise and/or editorial resources. Professional testing organizations go to great pains to produce items that are as "bulletproof" as possible, but for others offering moderate to high-stakes exams, this is not always feasible. It is likely the threat to exam score validity is even greater in such situations.

Organizations without sophisticated psychometric expertise would be wise to securely administer their exams to a sample of savvy test-takers in an effort to determine the extent to which the exam items are susceptible to guessing strategies. By asking examinees to provide the type of guessing strategy they used to respond to each item one can get a reasonable estimate of how much guessing is a threat to one's exam. Items deemed particularly problematic, or contaminated, could then be revised and administered on future exams. With proper equating, one could evaluate the effectiveness of the attempt to remove guessing contamination by Rasch analyzing the data and comparing the probability of success on the revised item relative to the item in its initial form. If the item's difficulty estimate increases after the revision, it is likely the revision was successful in removing much of the guessing contamination.

References

Rogers, H. J. (1999). *Guessing in multiple-choice tests*. In G. N. Masters and J. P. Keeves (Eds.). Advances in measurement in educational research and assessment. (pp. 23-42) Oxford, UK: Pergamon.

Kenneth D. Royal and Mari-Wells Hedgpeth
University of North Carolina at Chapel Hill

Suggestions for Improving AERA's Peer Review Process and Quality of Symposia

At the 2013 AERA annual meeting, those in attendance were asked to answer two questions regarding how future meetings could be improved. William Fisher provided the following suggestions:

What suggestions do you have for improving the quality of symposia at future AERA Annual Meetings?

Institute a peer-review rating system capable of supporting meaningful comparisons of proposal quality across SIGs and Divisions; one that (a) is built up from qualitative interviews and focus sessions, (b) is centered on a meaningful map of the construct, (c) includes items designed to represent the entire range of variation, (d) asks enough questions to drive down the error relative to the variation and obtain high reliability, (e) links all the raters to each other through the proposals they rate, (f) fits the data to a Rasch multifaceted model, (g) provides qualitative and quantitative formative feedback to the raters, proposal submitters, and summative feedback to the overall AERA membership.

For more info, see:

Andrich, D. (2010). Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika, 75*(2), 292-308.

Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences, 2d edition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Engelhard, G., Jr. (1997). Constructing rater and task banks for performance assessment. *Journal of Outcome Measurement, 1*(1), 19-33.

Engelhard, G., Jr. (2012). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York: Routledge Academic.

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

Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transactions, 7*(1), 283-284; [http://www.rasch.org/rmt/rmt71h.htm].

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Please provide any additional comments you may have.

It is very odd that a field so reliant on measurement for its most basic purposes is completely lacking in qualitatively meaningful, comparable quantities read off instruments traceable to common units. Everyone pays lip service to the importance of measurement, but almost no one goes to the trouble to seek out the state-of-the-art in instrument calibration or to implement highly advantageous and practical foundations for meaningful measurement in their research or teaching. Perhaps there is an important leadership opportunity here for someone able to bring these issues to the attention of the membership.

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

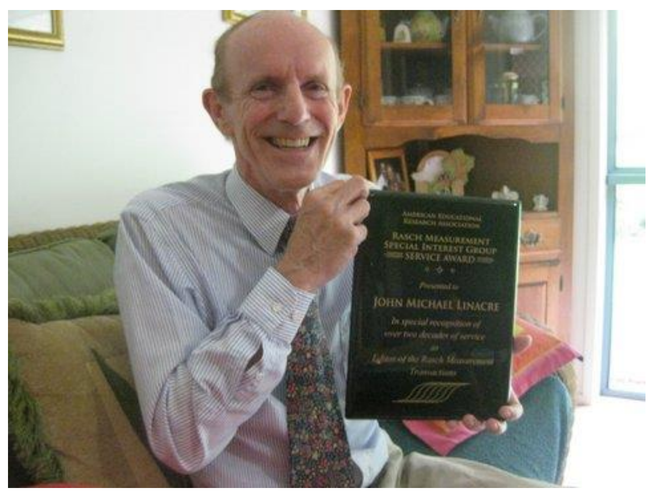
Rasch SIG Meeting Update from Chair

Greetings Rasch SIG colleagues,

I wanted to provide a brief review from the 2013 Rasch Measurement SIG Business Meeting, in case you were not able to attend. As you are likely aware this year's meeting was held in San Francisco, California at the Wyndham Parc 55 Hotel. I think "cozy, but comfortable" aptly describes the accommodations. Those in attendance were treated to good company and a nice assortment of refreshments. The meeting itself offered a chance to introduce current and outgoing SIG officers, report on membership and finances, summarize the 2013 Rasch contributions to the AERA conference, and provide an update on current business. As far as substance, membership is down slightly from last year, but perfectly in line with the trend over the past several years. Finances are effectively in line with what they have been. Of 19 proposals submitted, 13 were incorporated into 2 paper sessions and a roundtable. Outgoing Program Chair Daeryong Seo and remaining chair Kelly Bradley were thanked for this effort. An update was provided on the SIG bylaws, which to date is unchanged from my last update (expect to be contacted by AERA at some point to vote on final approval and implementation). Otherwise I mentioned that the awards proposed within the original bylaws needed to be reviewed separately by the AERA Executive Board and that hopefully this will be completed this June. Several discussion points were brought up during open floor. An announcement was made regarding status of plans for this year's International Objective Measurement Workshop in conjunction with AERA. Kirk Becker and I made a few announcements and appeals to include requests for volunteers to help out with various SIG activities and to note our collective intentions to consider possible future directions for the SIG. At that point in the meeting we welcomed our invited speaker Ed Wolfe who presented "Four Ways of Learning: Modeling Raters." Ed's presentation was engaging, relevant, and forward-thinking, which led to thoughtful discourse. Thanks again to Ed for this wonderful presentation. At that point, the meeting was adjourned.

One final note on the meeting, in line with action items from last year we were happy to be able to present a photograph of Mike Linacre having received his AERA Rasch SIG plaque in recognition of his editorial contributions to *Rasch Measurement Transactions*. He graciously agreed to allow us to share this. Thank you again, Mike!

Tim O'Neil
Pearson



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

RMT Editor Emeritus: John M. Linacre
Rasch SIG Chair: Tim O'Neil
Secretary: Kirk Becker
Program Chairs: Daeryong Seo & Kelly Bradley
Rasch SIG website: www.raschsig.org

Rasch SIG Service Opportunity:

WEBMASTER

The Rasch SIG would like to solicit a volunteer to serve as the webmaster for the Rasch SIG website. This individual will also be responsible for making *RMT* notes available on the web. Please email Editor\at/Rasch.org if you would like to volunteer, have questions or would like to know more about this service opportunity.

An Early Review of G. Rasch's Probabilistic Models

An Early Review of G. Rasch: *Probabilistic Models for Some Intelligence and Attainment Tests*, Danmarks Pædagogiske Institut, 1960. 184 pages. Danish kroner 20.

Georg Rasch is a mathematical statistician who does not only criticize the use of classical statistics for behavioral science problems due to his own (bad) experience; he goes on and formulates new models.

It has always been debated whether the results of a psychological experiment tell something about the persons tested or about the experiment. Rasch has ingeniously solved this problem replacing the set of stimuli with a parameter (called σ) and the set of personal factors also with a parameter (called θ). These parameters can be estimated; they may be as well one- as multidimensional.

(... A mathematical description of "Rasch's Measurement-model" ...)

We face here an important methodological improvement with analogies to other behavioral sciences than the psychological (children's reading ability has been the base for the model development). Certain extensions must be made, however, if the model should be used directly for measurement of other phenomena.

Erik Johnsen (1962) Mathematical Elements in Social Sciences, *Acta Sociologica*, 5(1), 170-171.

Ohio River Valley Objective Measurement Seminar (ORVOMS)

The third annual Ohio River Valley Objective Measurement Seminar (ORVOMS) was held on May 3, 2013 at the University of Kentucky. It was hosted by Kelly Bradley from the Department of Educational Policy Studies & Evaluation. The keynote speaker was William Boone from Miami University in OH.

This year's sessions included presentations on applied topics, such as using Rasch for pre/post assessments, instrument validation, evaluation studies, and interpreting Rasch output.

We thank everyone who attended and presented this year. We look forward to next year's seminar and hope that you will be able to participate. Initial planning for next year's meeting suggests Cincinnati, OH will be the meeting place. For information about upcoming events or to be placed on our mailing list please contact Melanie Lybarger (mlybarger \at/ theabfm.org).

Rasch-related Coming Events

June 19-21, 2013, Wed.-Fri. SIS 2013 Conference on Advances in Latent Variables: Methods, Models and Applications, Brescia, Italy, http://meetings.sis-statistica.org/index.php/sis2013/ALV

July 1-Nov. 30, 2013, Mon.-Sun. Online course: Introduction to Rasch Measurement theory (D. Andrich, RUMM), www.uwa.edu.au

July 5- Nov. 30, 2013, Fri.-Fri. Online workshop: Practical Rasch Measurement – Fruther Topics (E. Smith, Winsteps), www.statistics.com

Aug. 1-5, 2013, Thur.-Mon. TERA-PROMS Annual Meeting, Kaohsiung, Taiwan, tera.education.nsysu.edu.tw.

Aug. 9- Sept. 6, 2013, Fri.-Fri.. Online workshop: Many-Facet Rasch Measurement (E. Smith, Facets), www.statistics.com

Aug. 22, 2013, Thur. Symposium in honor of Svend Kreiner, Copenhagen, Denmark, biostat.ku.dk/kreinersymposium

Sept. 4-6, 2013, Wed.-Fri. IMEKO Symposium: Measurement Across Physical and Behavioural Sciences, Genoa, Italy, www.imeko-genoa-2013.it

Sept. 13-Oct. 11, 2013, Fri.-Fri. Online workshop: Rasch Applications in Clinical Assessment, Survey Research, and Educational Measurement (W. P. Fisher), www.statistics.com

Sept. 18-20, 2013, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Sept. 23-25, 2013, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK,

Sept. 26-27, 2013, Thur.-Fri. In-person workshop: Advanced Rasch (A. Tennant, RUMM), Leeds, UK

Call for Submissions

Research notes, news, commentaries, tutorials and other submissions in line with *RMT*'s mission are welcome for publication consideration. All submissions need to be short and concise (approximately 400 words with a table, or 500 words without a table or graphic). The next issue of *RMT* is targeted for Sept. 1, 2013, so please make your submission by Aug. 1, 2013 for full consideration. Please email Editor\at/Rasch.org with your submissions and/or ideas for future content.