RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 26 No. 2 | Autumn 2012 | ISSN 1051-0796

Acknowledgement: Mike Linacre's Service to RMT

For more than 20 years, Mike Linacre faithfully carried out RMT's editing duties. In acknowledgement of this extraordinary feat, some of Mike's friends and colleagues were invited to say a few words about his contributions to RMT. Here is what his peers had to say:

"When I first started to learn about Rasch analysis, Mike had just taken up the Editorship of RMT. Thus to me, RMT was, and has always been Mike. For more than 20 years, he has made a substantive commitment and, quite often, significant individual contributions to RMT. As editor he has encouraged innovative articles, and supported the methodological work that has often found a home within the pages of RMT. Indeed, some of the articles within RMT have been cited much more often than many who have found their way into the mainstream journals! That he has sustained this publication, year after year, issue after issue, is something that we should all be most grateful, and we owe him a great vote of thanks for his significant contribution to the growth of the application of the Rasch model." - Alan Tennant

"Mike Linacre combines the scholarly qualities of intelligence and erudition with friendliness and good will. He is hard-working and a diligent craftsman in his field. Respectful to all, Mike greets everyone with a big smile and honest concern. The field of measurement/statistics has been greatly blessed by his continuous efforts to inform, educate, and help. RMT is an outstanding product with Winsteps/Facets a boon to everyone. Thank you Mike and continued success." - Mark Stone

"Mike Linacre is among the rarest of academic colleagues and friends. He is first and foremost brilliant, with an encyclopedic command of his discipline. He is tireless in pursuit of new methods and new insights and he is always patient and gracious. The adjectives that pepper my conversations with colleagues about Mike and his contributions to our field are 'generosity' and 'humility'. Thanks Mike for 20+ years on RMT. You gave us our voice!" – Jack Stenner

"Mike Linacre's editorship of RMT has provided psychometricians, clinicians and other instrument users with timely, comprehensive and cogent information about Rasch measurement. He has given generously of his time and engaged contributors from around the world. His passion for objective measurement is reflected in high editorial quality of RMT. I look forward to reading each issue of RMT and maintaining my connection with Mike and other Rasch model users." – Allen Heinemann

"Mike Linacre developed and taught a number of Rasch courses at Statistics.com over the years, and gained a large and devoted following. Mike set the standard for online teaching, doing something I had not seen anyone do before -- he took an online classroom and made it as close to a warm, engaging in-person discussion as is humanly possible. Being a recognized expert in the subject helps, of course, but Mike also had a welcoming personality and an innate sense of how to guide students and bring them fully into the online classroom community - even in an asynchronous model of instruction. Via RMT, Mike has also demonstrated outstanding teaching through the years by providing clear and concise answers to tough measurement questions. I have learned much from Mike." – Peter Bruce

"I have had the privilege of knowing and working with Mike for many years. His stewardship of RMT made that publication one of the few that I can say I looked forward to receiving each month. RMT is like a special box of chocolates. Small morsels of knowledge, but each article worth savoring for its own unique contribution." – John A. Stahl

Table of Contents

"Mike Linacre and I met at the 1985 IOMW meetings in Judd Hall at the University of Chicago, and then again at the 1987 IOMW, also in Judd. We didn't really get acquainted though until we roomed together at the 1990 AERA in Boston, at Ben Wright's suggestion. We spent a lot of time together that week at the newly formed Rasch SIG sessions. I came to value greatly Mike's perspective over the course of that week. He displayed an amazing ability to focus on the key essentials of presentations that were—then as now—often obscured by irrelevant details and poor organization. As the years went by, Mike applied that ability in developing and maintaining the Facets and Winsteps software, in editing and writing for RMT, in consulting, and in teaching. His quick-turnaround responsiveness to questions and problems is legendary, as is his capacity for simplifying complex issues to clear fundamentals. We will all miss Mike's steady hand at the helm of RMT!" – William Fisher, Jr.

"The fact that we have had only one editor of Rasch Measurement Transactions (RMT) in the last 22 years speaks volumes about Mike's dedication to Rasch measurement and to his dedication to providing an accessible forum that allowed clear and useful professional interaction. Mike became editor beginning with Vol. 4, No. 2. I do not think that anyone envisioned the role RMT would play for the family of Rasch measurement practitioners when the SIG was first proposed in 1987. But 22 years later Mike has managed to amass an outstanding collection of research notes relating to all of the pressing issues in Rasch measurement, all the while making those notes direct and to the point, so that collection is accessible to the entire community of Rasch practitioners. These notes will remain a testament to Mike's dedication and foresight and are a valuable resource to all researchers in the field. Mike has always had an uncanny ability to spot the application of Rasch principals in a wide variety of scientific endeavors that have led to some very interesting and thought provoking issues of RMT." – Richard Smith

"Mike, at the news of your leaving the editor's position at RMT, I wish you the best in all you continue to do. I am glad that you are still continuing the other activities. I have personally gained much insight about the importance of Rasch due to your efforts in following Ben. Also, thanks for your valued and persistent teaching. Good wishes." – Carl Granger

"There are many remarkable things about Mike—the energy and enthusiasm he brings to his work, the pleasure and deep satisfaction he takes from it, his sense of humor, his kindness in helping others, and his lively intelligence. I've worked with Mike many times, mostly in the early part of his career in Rasch measurement, and it's always been a pleasure. I can't think of Mike without smiling. Mike took the Rasch computer programs of those days—MSCALE AND MSTEPS—which were written in FORTRAN mainly for batch runs on main frame computers and adapted them for running interactively on a PC using a Windows OS. Very soon he was producing the FACETS program for which he is now famous and backing it up with excellent statistical and technical documentation. Under Mike, the Rasch SIG newsletter itself became an outlet for much-needed statistical and technical documentation for Rasch measurement in general, but also accessible and open to all. Thank you so much, Mike, for your leadership and dedication and for the help you've been to me and others in Rasch measurement over the years. Thanks for editing the Rasch SIG newsletter. You deserve a rest! But I hope we will continue to see your gems in the newsletter." – Matt Schulz

"Looking back, I remember Mike and Ben would sneer at widely revered but arbitrary psychometric conventions (CTT, IRT, or whatever), while they flaunted the vastly superior Rasch models. Then they offered RMT as a safe haven, an intellectual niche, to anyone who would join them. In RMT, Mike and Ben cast a vision of what social science could be, namely, meaningful linear units, objective entities, and abstract measures. Mike's contributions to RMT are deeply embedded in this intellectual niche, where he shared inspirations and insights and solicited them from others. More importantly, Mike's influence on theory and practice continues to materialize in broader and more profound measurement applications. However, Mike's greatest RMT legacy will probably be establishing a self-sustaining intellectual force for advancing a new generation of Rasch scientists." - Nick Bezruczko

"After more than 20 years of service, Mike Linacre has stepped down from his post as Editor of *Rasch Measurement Transactions* (RMT). Kenneth Royal is taking over that task. Well, Mike might not look like the largest guy on the street, but those will be pretty big shoes to fill! While the rest of the Rasch measurement world goes (selfishly) on, self-absorbed in its own research, Mike was always beavering away, cobbling together the bits and pieces, notes, diagrams, citations and quotes that would become the next RMT. Always there, and always with some gem of insight to be discovered. My advice to newcomers is often: "Do a search on Rasch.org." For 20 years, Mike has been RMT; he has also been Rasch.org. Mike has always given us much more than we deserved. Thanks, Mike." – Trevor Bond

"Mike has edited and contributed to RMT for more than 20 years. During this time, RMT has progressed from being a newsletter for a relatively small group of Rasch aficionados to a key resource for measurement scholars in a variety of fields around the world who access it daily on the web. It is also important to acknowledge his important work in regards to two of the most widely used computer programs for invariant measurement: Winsteps and Facets. In the US alone, almost every statewide assessment program uses Winsteps for high stakes student

assessments. I can't think of anyone else who has played a larger role in translating the invariance principles of Rasch measurement theory into a user friendly set of computer programs that encourage and support sound psychometric practices. The Rasch measurement community has greatly benefited from Mike's dedication, commitment, and passion about measurement and I urge folks to go back and read (or re-read) earlier issues of RMT -- it provides a historical record of our accomplishments in measurement, and also contains quite a few gems regarding where we still need to do more work ... thank you, Mike!" – George Engelhard, Jr.

Mike, on behalf of all RMT readers, THANK YOU!

Transition of Editorship

Dear Rasch measurement enthusiasts,

Last year Mike Linacre announced he would be stepping down as the Editor of RMT after more than 20 years of service in this role. I have been appointed as Mike's successor and wanted to take an opportunity to introduce myself to you and briefly discuss my vision for RMT's future. First, I want to thank Mike for his service. He has certainly set the bar high for RMT, and in my view, is irreplaceable. Second, I am humbled by this opportunity to serve the measurement community and will do my best to carry the gauntlet forward. Like many of you, I have faithfully followed RMT for many years and have a great appreciation for this publication outlet. Unlike most SIG newsletters, RMT has demonstrated staying power. In fact, this year marks the 25th anniversary of RMT! In my view, RMT's longevity is due to numerous factors, but not the least of which is its contributors. Research notes that were published more than two decades ago are still very relevant today and continue to be cited in the research literature. The initial creators of RMT, Ben Wright and Richard Smith, intended for RMT articles to be "short, sharp, fast and useful". Obviously, this approach has been incredibly successful. My hope is to continue this tradition.

With regard to where I see RMT going, I have a few personal opinions, but would also be happy to hear suggestions from you. In my view, the greatest challenge ahead for the Rasch community is identifying new and innovative ways to apply Rasch models across a wide array of academic disciplines. Further, if we are to convince others of the utility of Rasch models and encourage their use, we must be able to demonstrate exactly how these models can be of benefit in everyday research and practice. As I solicit materials to publish in RMT, I will be keenly interested in new and innovative applications of Rasch measurement.

Of course, RMT has long been a wonderful outlet for not only research notes, but also news, commentary, and other musings. This tradition will continue. However, as we move forward I would like to add a few additional features. Because Rasch measurement has now witnessed several generations of Rasch enthusiasts with more gaining interest every day, it is important that this publication is tailored to an audience of varying experience levels, backgrounds, and disciplinary affiliations. As such, future issues will add features such as tutorials, an "Ask the Experts" series for particularly controversial topics, suggestions for "Teaching Rasch Measurement", and more. Although RMT is only slated to produce four issues per year, it is possible that additional issues focusing on special topics and themes will be published. I am open to any and all suggestions and would be happy to discuss any potential publication opportunities with you.

Finally, I would like to add that my previous statement regarding how it would be impossible for me to replace Mike Linacre is, in fact, an understatement. Few individuals on the planet have both the breadth and depth of knowledge that he possesses. However, with your help and continuous flow of quality submissions I am hopeful that we can continue to press forward with RMT in such a way that both Mike and the larger Rasch community will be proud.

Kenneth Royal

Mark Wilson's Psychometric Society Presidential Address

"Seeking a Balance Between the Statistical and Scientific Elements in Psychometrics", July 2012, Lincoln, NE

In his recent Presidential Address to the International Meeting of the Psychometric Society, Mark Wilson contrasted statistical and scientific themes in psychometrics in terms of the history of his own work, with larger goal of identifying scientific aspects in psychometrics that would distinguish it from statistical modeling. A paper based on this presentation will be forthcoming in *Psychometrika* in early 2013.

Early in his career, Mark developed the Saltus model of discontinuous development, which was and remains a highly innovative and effective guide to measuring cognitive growth. The work was done in isolation, however; it also is fairly complex, and it was not informed by input from anyone engaged with the substantive practicalities of research in cognitive development. Thus, Mark pointed out that his publications in this area are rarely cited and his model has had little or no impact.

This situation is quite different from more recent work Mark has been doing, in which engagement with substantive content experts is an essential ingredient. Now, the models and construct theories implicitly or explicitly included in curricular outcomes assessments are articulated, developed, and applied in collaboration with experts in the substantive area. As Mark pointed out, there is no dumbing down of model complexity necessary in this context, as many projects naturally entail multiple constructs manifest at multiple levels of organization and/or with multiple facets, and which are assessed via many different types of items or performance rating schemes, some of which may involve testlets or item bundles and their local dependencies.

Mark illustrated one of these new collaborations as an example. A middle and high school statistics and data modeling curriculum developed at Vanderbilt University involves a number of separate but interdependent strands. To understand just what was intended for the assessment and to formulate a plan adequate to the needs of both instruction and accountability:

A. detailed theoretical maps of each construct's levels and sublevels were laid out;
B. items were designed to express each level of each construct, and with an eye to the interrelations between those constructs;
C. the scoring of the items, distractors and mistaken responses was set up to inform individualized instructional applications;
D. the measurement model appropriate to the overall assessment system is then applied to a pilot data set; and
E. the new information on the system performance is then used to revise the construct map(s), the item design, the outcome space, and the model for ongoing applications.

Though Wilson did not mention it, readers of his 2005 *Constructing Measures* text will recognize here the four phases of that book's systematic assessment methodology. And completing a round or two through the process certainly sets the stage for iterating through it once again, with the intention of taking the construct sublevels to a new level of specificity capable of affording predictive control over item design and scoring, as is suggested in his 2004 book with DeBoeck on explanatory models.

In conclusion, Wilson raised again the question of just how psychometrics is to be more than a specialized branch of statistics if it does not capitalize on the practical opportunities for measurement it has created for itself in education, psychology, and other fields. There seems to be great potential for integrating qualitative substantive theory and practice with quantitative methods and modeling. Perhaps a recognizable new paradigm is now in the process of forming.

William P. Fisher, Jr.

Rasch Measurement on LinkedIn

The Rasch Measurement group on LinkedIn is rapidly growing. As of January 1, 2012, the group had 285 members. Nearly eight months later, the group had approximately 500 members. If you would like to join us on LinkedIn, please visit the following url:

http://www.linkedin.com/groups?home=&gid=111741

RMT is also on LinkedIn!

Please consider joining the RMT group on LinkedIn as well. This group will serve as a forum to discuss Rasch measurement, as well as solicit ideas and opinions for future content to appear in RMT. To join the group, please visit:

http://www.linkedin.com/groups?gid=4581185&trk=myg_u

Rasch in England

When George Rasch ignored the existing approaches to analyzing test data and published his novel approach in 1960 he made a major psychometric breakthrough opening up possibilities which even he did not foresee. The significant advance showed how sample free instruments could be created which measured on an equal interval scale and how person abilities could be estimated regardless of the items used.

The ideas were championed and extended by Ben Wright and his associates at the Measurement, Evaluation, Statistics and Assessment (MESA) Unit at the University of Chicago and picked up by others including Bruce Choppin in the UK. It was Choppin's vision to use the theoretical advances to overcome problems associated with monitoring standards over time. This would involve creating a database of items with known difficulty and, using, matrix sampling, create a national system which could be used to track standards whilst providing others with the means to probe more deeply (Choppin 1969, 1981).

This vision was ahead of its time and because of objections brought forward by skeptics, came to an abrupt halt in the late 1980. The health of educational testing in England was damaged and, arguably, has never fully recovered. As Mike Linacre (1995) wrote

> *'Under Choppin's supervision British psychometrics could have led the world (to the great benefit of British students, teachers, and policy makers). Instead the entrenched interests condemned Britain to a 60 year regression'*

The fundamental issues which were used to challenge the innovative approach have been discussed at length in the literature and, with hindsight, one can see the misunderstandings and lost opportunities which, at the time, were only seen by the farsighted. The issues included unidimensionality, the constancy and variation of item difficulties and the fit of people to the model.

Meanwhile workers such as Andrich and Linacre have taken the field to a different level and the influence of Rasch can be seen throughout the psychometrics world. Thomas Kuhn (1962) would have recognized the paradigm shift that is still taking place.

Peter Tymms – Durham University (UK)
Panayiotis Panayides – Lyceum of Polemidia (Cyprus)

References

Choppin, B. H. (1969) An item bank using sample-free calibration, in: R. Wood & L. S. Skurnik (Eds) *Item banking* (Slough, National Foundation for Educational Research).

Choppin, B. (1981) Is education getting better? BERA Presidential Address, University College, Cardiff, 1980, *British Educational Research Journal*, 7(1), 1–15.

Kuhn, T. (1962) *The structure of scientific revolutions* (Chicago, IL, University of Chicago Press).

Linacre, J. M. (1995) Bruce Choppin: visionary. Available online at: http://www.rasch.org/rmt/rmt84e.htm.

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests* (reprinted in 1980 with a foreword and afterword by Benjamin D. Wright) (Chicago, IL, MESA Press).

Postscript

A fuller account of the story recounted above and the issues surrounding it, including responses to the major criticisms of the model, can be found in:

Panayides, P., Robinson, C. and Tymms, P. (2009) 'The assessment revolution that has passed England by: Rasch measurement *British Educational Research Journal 36*, 4, 611-626

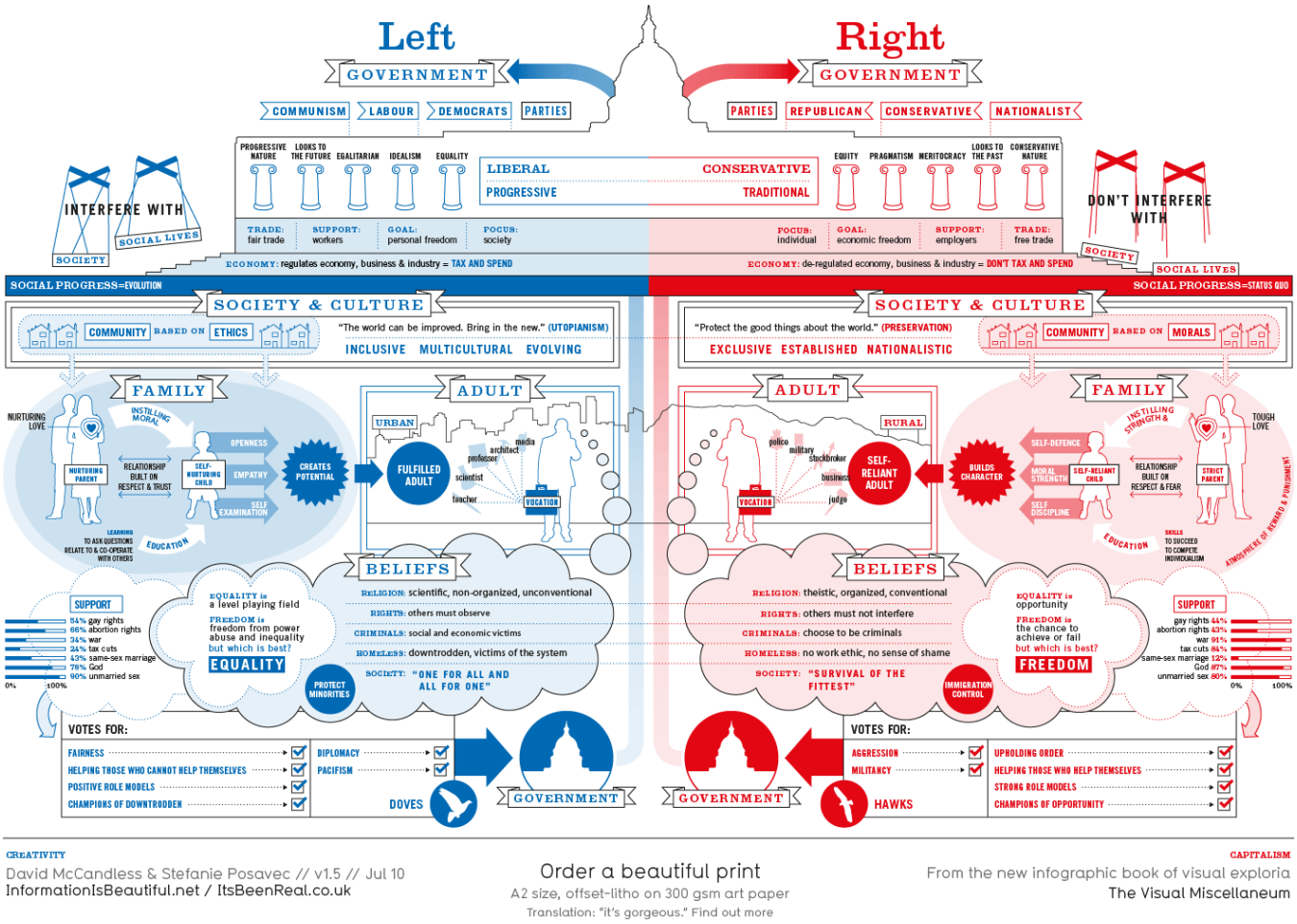
Kaggle Competitions

Rasch measurement enthusiasts may be interested in participating in one or more of Kaggle's open competitions. The Kaggle website features a steady stream of data science competitions. Some notable competitions currently underway include the "Heritage Health Prize" which asks participants to identify patients who will be admitted to a hospital within the next year using historical claims data. A prize of $3 million is available to the winner!

Other competitions include Careerbuilder's "Job Recommendation Engine Challenge", where participants are asked to predict which jobs users will apply based on their previous applications, demographic information, and work history. A $20,000 prize is available for this competition.

Visit www.kaggle.com for a comprehensive list of current competitions.

An Excellent Visualization of Political Spectrum Data in the United States



Created by David McCandless and Stefanie Posavec. Available at: http://www.informationisbeautiful.net/2009/left-vs-right/

Measuring Liberal/Conservative Voting Tendencies among U.S. Senators

Each year, the *National Journal* examines roll-call voting records for United States Congress, and groups topical voting records into three categories: economic, social, or foreign. This Rasch analysis attempted to measure liberal versus conservative voting tendencies based on 2011 voting records. Considering the House of Representatives consists of 435 members and the Senate consists of 100 members, for convenience, only U.S. Senators were investigated.

In total, 235 Senate voting records were registered in 2011. Many of these votes involved noncontroversial issues or topics that did not invoke ideological distinctions and thus, were removed from the dataset. In total, 97 votes appeared to fall along party lines and were selected as the basis for this analysis. These votes were parsed into the three aforementioned categories.

In order to analyze the data, it was necessary to first create two distinct data sets, one for senators who identified themselves as "Democrat" (n=51) and one for senators who identified themselves as "Republican" (n=47). Measures of conservative/liberal voting behaviors were discerned relative to members of one's own political party. This data parsing allowed two questions to be investigated: Among all Republican senators, who is the most/least likely to provide a conservative vote? And among all Democratic senators, who is the most/least likely to provide a liberal vote?

The *National Journal* data set provides counts of liberal versus conservative votes for each of the topical categories. This information alone is not useful for a Rasch analysis. However, with a rather novel recoding schema a useful data set can be prepared. For example, suppose a Republican senator placed 79 conservative votes and 16 liberal votes on foreign issues. This senator would provide a conservative vote on foreign issues at a ratio of about 5:1. Thus, simply assigning a value of 5 can serve as a useful proxy for the magnitude of conservative voting on topics of this nature. This recoding schema was repeated for every senator in both data sets until a useful

data set was constructed. All data were recoded into values ranging from 1 to 9, with ratios exceeding 9 being truncated to 9.

Finally, person calibrations were produced for both Republicans and Democrats. Logits values were then rescaled onto a continuum ranging from 1-10 for easy interpretation. Results are presented in Tables 1 and 2.

Political scientists often refer to a political spectrum that ranges from the far liberal left to the far conservative right. Where one's views fall along this spectrum likely will determine the extent to which common ground can be established between individuals from opposing political parties. As demonstrated in this analysis, Rasch measurement can help empirically present one's views along this political spectrum. Senators with the highest measures are typically more polarizing in their views, whereas persons with lower measures are more likely to entertain views from the opposing party. Information gleaned from analyses such as this one can be useful for: identifying individuals with a voting record that likely will (or will not) resonate well with voters; identifying the extent to which legislators' votes are consistent with their political platforms; comparing the voting records/political views of legislators from the same state; identifying which individuals are likely to filibuster a bill if given the opportunity; predicting how various legislators will vote on a highly partisan issue; predicting the productivity of various congressional subcommittees based on its panel of members, etc.

Table 1. *Conservative Voting Tendencies for Republican Senators*

Senator	State	Measure	SE	Senator	State	Measure	SE
Coburn	OK	10.00	1.88	Heller	NV	7.81	0.36
Johnson	WI	10.00	1.88	Moran	KS	7.81	0.36
Crapo	ID	8.60	0.43	Paul	KY	7.81	0.36
Risch	ID	8.60	0.43	Shelby	AL	7.81	0.36
Inhofe	OK	8.49	0.39	Coats	IN	7.73	0.38
Barrasso	WY	8.40	0.36	Hutchison	TX	7.73	0.38
Burr	NC	8.40	0.36	Chambliss	GA	7.63	0.41
DeMint	SC	8.40	0.36	Isakson	GA	7.63	0.41
Enzi	WY	8.40	0.36	Roberts	KS	7.50	0.46
McConnell	KY	8.40	0.36	Hoeven	ND	7.34	0.54
Vitter	LA	8.40	0.36	Wicker	MS	7.34	0.54
Cornyn	TX	8.25	0.34	Alexander	TN	7.11	0.67
McCain	AZ	8.25	0.34	Portman	OH	7.11	0.67
Sessions	AL	8.25	0.34	Johanns	NE	6.01	1.21
Hatch	UT	8.18	0.33	Lugar	IN	6.01	1.21
Kyl	AZ	8.18	0.33	Blunt	MO	4.85	1.45
Rubio	FL	8.11	0.33	Graham	SC	4.85	1.45
Ayotte	NH	7.97	0.33	Kirk	IL	4.85	1.45
Grassley	IA	7.97	0.33	Cochran	MS	3.21	1.80
Lee	UT	7.89	0.34	Murkowski	AK	3.21	1.80
Thune	SD	7.89	0.34	Brown	MA	1.00	2.97
Toomey	PA	7.89	0.34	Collins	ME	1.00	2.97
Boozman	AR	7.81	0.36	Snowe	ME	1.00	2.97
Corker	TN	7.81	0.36				

Table 2. *Liberal Voting Tendencies for Democratic Senators*

Senator	State	Measure	SE	Senator	State	Measure	SE
Akaka	HI	10.00	2.27	Casey	PA	7.36	0.67
Durbin	IL	10.00	2.27	Johnson	SD	7.36	0.67
Gillibrand	NY	10.00	2.27	Udall	CO	7.36	0.67
Merkley	OR	10.00	2.27	Blumenthal	CT	7.09	0.80
Boxer	CA	8.83	0.74	Carper	DE	7.09	0.80
Brown	OH	8.83	0.74	Hagan	NC	7.09	0.80
Cardin	MD	8.83	0.74	Klobuchar	MN	7.09	0.80
Harkin	IA	8.83	0.74	Menendez	NJ	7.09	0.80
Mikulski	MD	8.83	0.74	Tester	MT	7.09	0.80
Udall	NM	8.83	0.74	Warner	VA	7.09	0.80
Leahy	VT	8.31	0.51	Begich	AK	6.70	0.96
Franken	MN	8.19	0.49	Bennet	CO	6.70	0.96
Lautenberg	NJ	8.19	0.49	Cantwell	WA	6.70	0.96
Rockefeller	WV	8.19	0.49	Conrad	ND	6.70	0.96
Feinstein	CA	8.07	0.48	Kohl	NH	6.70	0.96
Reed	RI	8.07	0.48	Shaheen	VA.	6.70	0.96
Whitehouse	RI	8.07	0.48	Webb	LA	6.25	0.91
Inouye	HI	7.96	0.49	Landrieu	WV.	5.90	0.76
Murray	WA	7.96	0.49	Manchin	MO	5.90	0.76
Reid	NV	7.96	0.49	McCaskill	FL.	5.90	0.76
Schumer	NY	7.96	0.49	Nelson	MT	5.90	0.76
Stabenow	MI	7.96	0.49	Baucus	MI	4.72	0.84
Wyden	OR	7.96	0.49	Levin	NE	4.26	1.16
Coons	DE	7.84	0.50	Nelson	AR.	1.00	3.53
Kerry	MA	7.55	0.58	Pryor	NH	1.00	3.53
Bingaman	NM	7.36	0.67				

Kenneth D. Royal
University of Kentucky

IRT-based Automated Test Assembly using Constraint Logic Programming (CLP)

It is supposed that an item pool consisting of a large number of items with pre-calibrated statistics (such as item difficulty in logit and item fit - infit/outfit statistics) has been established. With such a provision, users may want to extract a small number of items from the pool to set up a test fulfilling a number of criteria, namely:(i) specific test length; (ii) specific numbers of MC and non-MC items; (iii) number of items of a category being greater than a particular minimum value; (iv) items' difficulty values and fit values falling in particular ranges; (v) specific items that should be included or excluded; and (vi) the cumulative item information at a certain logit value (say 0) being greater than a particular minimum value. This kind of IRT-based automated test assembly has been well investigated under the framework of Integer Programming. If the item pool consists of N items, the test is modeled as a vector of decision variables of size N. Each decision variable is either 0 or 1 standing for whether the corresponding item is excluded or included in the test. The above mentioned criteria could be defined as a set of linear (in)equalities, in terms of decision variables.

In this summary notes, it is proposed to apply a novel approach to IRT-based automated test assembly. In this approach, Constraint Logic Programming (CLP) is used, instead of Integer Programming. In CLP, a test is represented as a set of item IDs. Constraints could be set up on the cardinality and membership of the set. Constraint propagation techniques and backtracking search are then employed to generate a feasible solution satisfying all the constraints. A number of advantages of using CLP are stated below:

(i) Representation of a test as a set of item IDs is much more compact, as compared with that of using a vector of 0-1 elements of size N, where N could be up to thousands if the item pool is large. This implies computational efficiency.
(ii) Representation of a test as a set of item IDs is more natural and intuitive. This facilitates the development and interpretation of the program.
(iii) Logic programming is declarative in nature. Thus the program code of CLP (which could be simply regarded as a specification of requirements) is much easier to be understood, as compared with the mathematical formulation of Integer Programming.
(iv) In CLP, non-linear constraints could be handled; while in Integer Programming only linear constraints could be entertained.
(v) CLP provides a great flexibility in formulating constraints. For example, symbolic constraints (i.e., constraints consisting of non-numerical values) could be built.
(vi) In Integer Programming, an objective function for minimization (or maximization) must be established for running the algorithm; while in CLP there is no such a need.
(vii) Based on backtracking search, more than one feasible solution could be generated as alternatives.

An illustrative example

In the following, we illustrate the CLP approach using a simulated item pool with 165 items. The freeware Eclipse Constraint Logic Programming (available from http://eclipseclp.org/) is adopted. The items and the corresponding summary statistics are represented using the following programming clauses:

```
% ID, Type, Difficulty, FitValue, ItemInfo, Category
simData(1,1,3.73,1.69,0.02,4).
simData(2,1,1.74,0.43,0.12,4).
…..
```

We aim to set up a test of length 30, which is specified using the following program clauses:

```
intset(TestSet, 1, 165), %TestSet is a set of item IDs from 1 to 165 #(TestSet, 30). % of size equal to 30
```

Next, we specify the number of MC items in the test being equal to 10 using the following program clauses:

```
findall(Item, simData(Item,1,_,_,_,_), List1),%find all IDs of MCs ic_sets:intersection(TestSet, List1, Res1), %Res1 is TestSet ∩ List1 #(Res1, C1),C1 #= 10. %Cardinality of Res1 is 10
```

In logic programming, a very handy utility, findall, is provided to pick up all solutions of a query together in a list. In the above programming clauses, findall(Item, simData(Item,1,_,_,_,_), List1) is used to pick up all item IDs of MC type (i.e., Type=1) and store them in the List, List1.

Suppose there are 4 categories. The minimum numbers of these categories are: [2,2,2,4]. As an illustration, the following program clauses are used to set up the constraint for the first category.

```
findall(Item, simData(Item,_,_,_,_,1), List), %find all IDs of cat1 ic_sets:intersection(TestSet, List, Res), %Res is TestSet ∩ List #(Res, C),  2 #=< C. %Cardinality of Res ≥ 2
```

We only want to include only those items whose difficulty values, say falling in [-1.5. 2] and fit values in [0.1, 1.6]. This can be achieved using the following program clauses:

```
findall(Item, satisfy(Item), List), %find all IDs fulfilling criteria
List includes TestSet. % TestSet is a subset of List

%The program clauses specify the criteria on difficulty and fit satisfy(Item):-
    simData(Item,_,Diff,Fit,_,_),
```

2 >= Diff,Diff >= -1.5,
1.6 >= Fit, Fit >= 0.1.

We may want to exclude and include some specific items. As an illustration, suppose that we want to exclude item 110, and include item 129 and 134. The corresponding program clauses are as follows:

TestSet disjoint [110],% TestSet ∩ [110]= ∅
TestSet includes [129,134]. % [129, 134] is a subset of TestSet

The above program clauses should be quite intuitive and self-explanatory, even for those who are not familiar with logic programming. Finally, we have to set the constraint on the cumulative item information of the test. Suppose that we want the cumulative item information at the 0 logit value being greater than or equal to 4.0. We could use the following program clauses:

weight(TestSet, Weights, ExpectInfo),
ExpectInfo #>= 4000 .

In the above program clause, weight is a built-in utility, which allows the association of a weight value (stored in the array Weights) to each set element. In our example, the array Weights stores the information value of each item one by one. Since only integral values are allowed to be stored, we transform the scale by multiplying with 1000, and the constraint becomes ExpectInfo #>= 4000, where ExpectInfo is the cumulative item information of the items stored in TestSet.

After setting up all the constraints, we request the system to start the searching process by using the following program clause: insetdomain(TestSet,_,_,_). The insetdomain is used to enumerate all ground instantiations of a set variable, and during the searching process constraint propagation techniques are used to reduce possible alternatives (domain reduction) in order to improve the efficiency.

After loading the program and running the query, the outputs from Eclipse CLP platform are captured below:

?- query(TestSet, ExpectInfo).
TestSet = [3, 8, 11, 12, 16, 17, 21, 29, 31, 32, 34, 35, 40, 44, 51, 55, 56, 69, 129, ...]
ExpectInfo = 4.011
Yes (0.03s cpu, solution 1, maybe more)

Category 1: 4
Category 2: 5
Category 3: 7
Category 4: 4
3
8
11
12
16
17
21
29
31
32
34
35
40
44
51
55
56
69
129
134
----------the end of the test---------

A feasible solution is generated within a very short period of time (0.03s cpu). More solutions could be obtained by clicking the tab More provided in the platform.

Fung Tze-ho
Hong Kong Examinations and Assessment Authority

Terms that have different meanings for scientists and the public

Scientific term	Public meaning	Better choice
enhance	improve	intensify, increase
aerosol	spray can	tiny atmospheric particle
positive trend	good trend	upward trend
positive feedback	good response, praise	vicious cycle, self-reinforcing cycle
theory	hunch, speculation	scientific understanding
uncertainty	ignorance	range
error	mistake, wrong, incorrect	difference from exact true number
bias	distortion, political motive	offset from an observation
sign	indication, astrological sign	plus or minus sign
values	ethics, monetary value	numbers, quantity
manipulation	illicit tampering	scientific data processing
scheme	devious plot	systematic plan
anomaly	abnormal occurrence	change from long-term average

Reprinted with permission from Somerville, R. C., & Hassol, S. J. (2011). Communicating the science of climate change. *Physics Today*, *64*(10), pp. 48-53. Copyright 2011, American Institute of Physics.

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

SIG Chair: Tim O'Neil
Secretary: Kirk Becker
Program Chairs: Daeryong Seo & Kelly Bradley
SIG website: www.raschsig.org

Rasch Item-bank Open Source Project

Purpose

The purpose of the Rasch Item-bank Project is to create an objective and highly reliable numeracy assessment tool, which is free and available to children all over the world.

There is an abundance of commercial education software on the web and there are many free games or game-like education applications, but there are few serious measurement tools. The idea of this project is to create a self-assessment tool enabling children in sub-Saharan Africa or the mountains of Peru to be measured on the same scale as children from central London or Manhattan.

Theory

The project is inspired by the chapters in Probabilistic Models (Rasch 1960), which deal with reading rates (Chapters 3 and 9). The software applies the arguments in those chapters to use the scoring rates directly to estimate item difficulty and student ability.

Scoring rates are defined as the reciprocal of the time taken (in fractions of a minute) to complete each item. Timing item responses invisibly on each item averts the theoretical and practical concerns expressed in the literature (Ebel, 1979, Meijer, 1993, Halkitis, Jones & Pradhan, 1996) over tightly timed pencil and paper tests.

Research by the project owner (Hippisley & Douglas, 1998) indicates that scoring rates on number skill tests make a highly reliable metric. Scoring rates on individual items correlate well with each other, with the mean rates on whole sessions, and with mean rates on other sessions by the same student.

Method

The method of the software is:

- To serve up test items from a bank of items stored on a file server;
- To record a dichotomous item score on each item;
- To record the scoring rate (reciprocal of the time taken) on each item;
- To adjust the difficulty of items served in response to the ability implied by the results on items previously served;
- To record a unique session parameter for each student test session;
- To return the items scores, scoring rates and session parameters to the server.

By default, for the purpose of Rasch analysis the session replaces the student. This enables anonymous data to be collected and analyzed, without contravening the data protection and privacy laws, which apply in many countries. In cases where permission has been granted to record data against students, actual students can be used in the analysis.

Roles of the project owner

- To create source code for a demonstration web based application (done);
- To make the demonstration application available on a public server (done);
- To analyze data from the application and refine item grouping.
- To raise awareness of the project and the application.

Possible roles of future participants

- Programmers may create source code for the generation of additional test items;
- Researchers may use the published application to carry out addition research, perhaps further researching the reliability of the application, or correlating results with other instruments;
- Teachers may simply use the application for their own diagnostic purposes.

Resources

- Project URL: http://java.net/projects/rasch-itembank/
- Demonstration application URL: http://softway.org/PP120101/AMJApp.php.

Project owner

Jonathan Hippisley, PhD
Email: jhipp\at/softway.org

References

Ebel, R.L. (1979) *Essentials of Educational Measurement* Prentice Hall, New Jersey.

Halkitis, P.N., Jones, P.J., Pradhan, J. (1996) *Estimating Testing Time: The Effects of Item Characteristics on Response Latency* ED397119

Hippisley, J., Douglas, G. (1998) The reliability and validity of an interactive arithmetic test *British Journal of Education Technology*, 29(4), 303-320

Meijer, J. (1993) Learning potential, personality characteristics and test performance *Learning Potential Assessment: Theoretical methodological and practical issues* (Hamers JHM, Sijtsma K, Ruijssenaars JM Eds) Swets and Zeitlinger, Amsterdam, NL 341-362

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests* University of Chicago Press, Chicago

Announcing the Ideas in Testing Research Seminar, October 12, 2012

Graduate students, researchers and testing professionals are cordially invited to attend the first Illinois Institute of Technology (IIT) Ideas in Testing Research Seminar.

Where: McCloska Ballroom, McCormick Tribune Campus Center, 3201 S State Street, Chicago, IL 60616
When: 10:00AM to 5:00PM on October 12, 2012
Who: Graduate students, researchers and testing professionals

This is an excellent opportunity for members of the Midwest testing community to come together to discuss ideas, share expertise and develop professional relationships. The seminar will include presentations of research in progress, discussion of current trends in testing and an opportunity to meet with colleagues in a relaxed and friendly environment.

Lunch will be provided and there is no cost for registration however a reservation is requested. Reservations for attendance may be made by emailing Kirk Becker (kirk.becker\at/pearson.com) by September 30, 2012. Chicago possesses a rich variety of wonderful restaurants and there will be an opportunity to join your colleagues after the seminar for drinks or dinner.

Submissions to the seminar are encouraged. The range of acceptable topics is broad and includes but is not limited to:

- CAT
- Differential item functioning/measurement equivalence
- diagnostic testing
- automated item generation
- assessment in personality, work or clinical settings
- natural language processing (NLP) and computational linguistics (CL) applied to psychometric problems
- Simulations and games-based assessment
- Evidence centered design
- Bayes nets, naïve Bayes, MIRT, Rasch, or other advanced models
- Techniques for estimating model parameters

Potential presenters should submit a brief summary of the research to Alan Mead (mead\at/iit.edu) and Kirk Becker by September 1, 2012. Summaries should be no longer than 500 words excluding tables, figures, exhibits, or references and should describe the potential presentation.

Questions about the seminar may be directed to Alan (mead\at/iit.edu) or Kirk (kirk.becker\at/pearson.com) at the emails given above. We hope you will join us.

Rasch-related Coming Events

Sept. 5-7, 2012, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Sept. 10-12, 2012, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK,

Sept. 13-14, 2012, Thurs.-Fri. In-person workshop: Advanced Rasch (A. Tennant, RUMM), Leeds, UK,

Sept. 14 – Oct. 14, 2012, Fri.-Sun. On-line course: Rasch Applications in Clinical Assessment, Survey Research, and Educational Measurement (W.P. Fisher, Winsteps), www.statistics.com/raschapps

Oct. 1, 2012, Sun. Online Degree Programs: Application deadline: Measurement, Evaluation, Statistics and Assessment (E.V. Smith), MESA, University of Illinois at Chicago. http://www.uic.edu/scs/education/degree/online-mesa/

Nov. 2-3, 2012, Fri. - Sat. Workshop on "An Introduction to Rasch Measurement: Theory and Applications". (E.V. Smith & R.M. Smith), Chicago, IL.

Nov. 30, 2012, Fri. Workshop on "Rasch Models in Business Administration. Analysis Tools in Times of Crisis", Tenerife, Spain, IUDE, Universidad de La Laguna.

Dec. 5-7, 2012, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK,

Dec. 10-12, 2012, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK,

March 25-27, 2013, Wed.-Fri.. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK,

Apr. 27 – May 1, 2013, Sat.-Wed. AERA Annual Meeting, San Francisco, CA, www.aera.net

May 15-17, 2013, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK,

May 20-22, 2013, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK.

Sept. 18-20, 2013, Wed.-Fri.. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK,

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

Rasch Q&A

Question: How would a unidimensional model deal with something like the SAT, PISA, or any instrument that has strong multidimensional aspects? My response would be that we should separate the subject matter, but established instruments like these will likely not be changed in order to conform to unidimensional requirements.

Answer: The original thermometers, in the earlier 1600s, were multidimensional. They were open-ended glass tubes that combined the measurement of heat and of atmospheric pressure. Imagine if physicists around 1650 had said "our established open-tube thermometers will likely not be changed in order to conform with unidimensional (heat or atmospheric pressure) requirements."

Now, compare progress in the last 100 years for the social and physical sciences. Which science has the more effective methodology? Could social science research have done worse if it had imposed the unidimensional rigor of physical science upon itself?

SAT scores are treated as though they are unidimensional, so the SAT should be designed that way. Of course, "unidimensional" depends somewhat on the context. In the context of learning difficulties, "subtraction" is a dimension. In the context of academic achievement, "math" is a dimension. It is the same in physics; in some situations, "heat transmission by radiation", "heat transmission by conduction" and "heat transmission by convection" are different dimensions, but, for most purposes, "heat" is one dimension.

John M. Linacre

Consider joining a Rasch Listserv

ACER Listserv

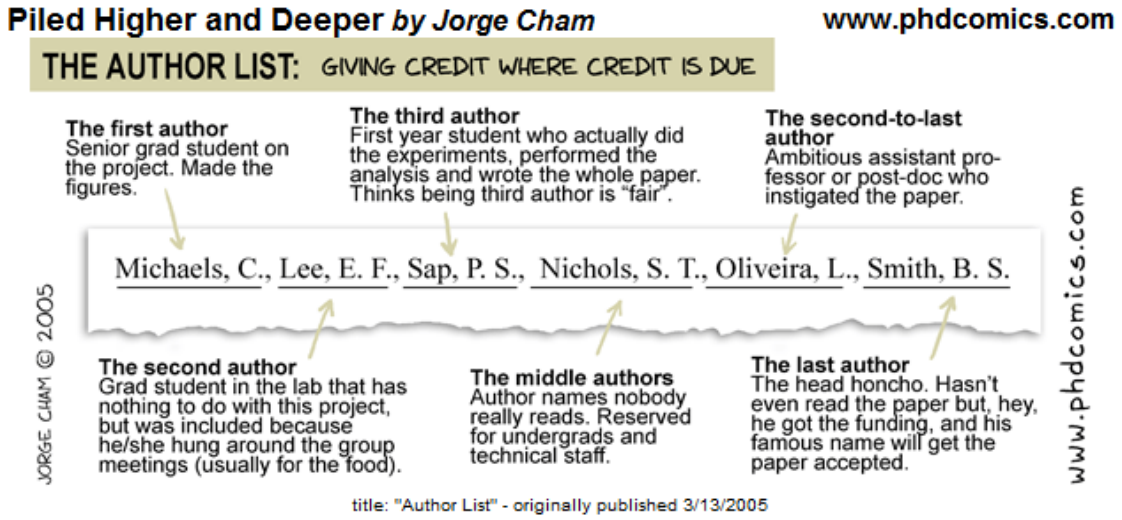
To join the Rasch discussion LISTSERV, send an e-mail with text *"subscribe rasch"* to: mailserv\at/acer.edu.au

Matilda Bay Club (MBC)

The MBC maintains a multidisciplinary discussion list dedicated to scientific measurement based on the principles of Rasch Measurement. Please visit MBC for more information.

Call for Submissions

Research notes, news, 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 January 1, 2013, so please make your submission by December 1, 2012 for full consideration. Please email Kenneth Royal at: Editor\at/Rasch.org with your submissions and/or ideas for future content.



"Piled Higher and Deeper" by Jorge Cham www.phdcomics.com