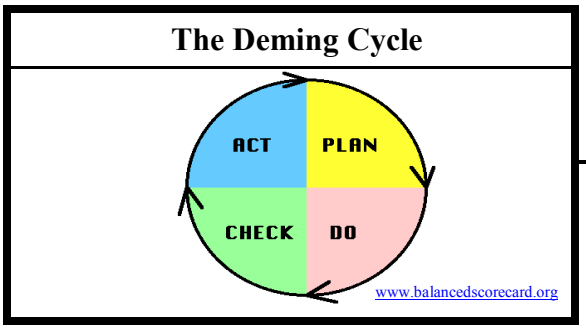


RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 26 No. 1 | Summer 2012 | ISSN 1051-0796

On Temperature

Figure 1: Bruegel's ***Temperance***. From: employees.oneonta.edu/farberas/

The 1560 print of *Temperance* (Figure above) by Pieter Bruegel, the Elder, illustrates *"pantometry"* (geometrical measurement) in action. The upper right section of this print portrays practical applications of the mathematical sciences and measurement. These scenes illustrate quantification attempts across many aspects of measuring; using a divider, square, plumb line, visual sighting, aspects of velocity/distance with cannons or crossbow together with disputation also serving a prominent role. (For more about this picture, see Crosby, 1997.)

Quantification and visualization go hand in hand with observations by providing the key to understanding measurement. Arithmetic, geometry, and trigonometry share with writing and music the pursuit of uniform quanta. Writing and music are linear events no less embodying measurement than any other area of science.

Table of Contents

Bruegel captured more than just the historical scene, he pictures the essence of metrology – a continuous search for units with generality.

Application and usefulness of units requires that all measures (and units) possess *sensus communis* or "common sense" as Kant (1917) expressed it. Kant meant that communication among peoples is not possible without a "common sense" operating. Visualizing measurement is applying common sense by the use of pictures, graphs, maps, etc. This approach is the key to success in communication, utility and generality (Stone, Wright & Stenner, 1999).

Measurement is always made by means of an analogy. Hans Vahinger (1924) wrote,

> All cognition is the apperception of one thing through another … we are always dealing with an analogy and we cannot imagine how otherwise existence can be understood … all knowledge can only be analogical. (p. 29)

Common examples from the past for measuring time include the tolling of bells, sundials and water-clocks. Today we have digital watches and atomic clocks for measuring time with greater accuracy. "Time passes," we say. "Time marches on," and when it does we record the duration in terms of length. There is no "time," only duration. Length is the analogy for duration. A theory of time as duration is transformed by analogy from a variable of length and made manifest using natural occurrences such as the sun, moon and stars, and artificial devices as mentioned earlier.

Robert Oppenheimer (1955) in his address to the American Psychological Association entitled *Analogy in Science* said:

> Whether or not we talk of discovery or of invention, analogy is inevitable in human thought, because we come to new things in science with what equipment we have, which is how we have learned to think, and above all how we have learned to think about the relatedness of things. We cannot, coming into something new, deal with it except on the basis of the familiar and the old fashioned. … We cannot learn to be surprised or astonished at something unless we have a view of how it ought to be; and that view is almost certainly an analogy. (p. 129-130)

Rasch (1961) addressed this problem with a theory, a class of models and specific data examples. His goal was "replacing qualitative observations by quantitative parameters" (p. 331).

Consider temperature and its common measurement. Temperature for most of us means the heat or cold we experience in our environment. In laboratories it is more rigorously studied, but in day-to-day life as well as in scientific laboratories, temperature requires some analogous method by which to make measures. A thermometer commonly uses an expansion tube of mercury to accomplish this task. Water, alcohol among other elements were investigated in arriving at the choice for mercury. Variations abound on the way to utility.

Consider the common indoor/outdoor device, the thermometer often showing both Celsius and Fahrenheit: shown in Figure 2.

For practical purposes the thermometer is simply an "expansion tube" of mercury. The elevation (length) of mercury in the tube is analogous to temperature. This elevation is made utilitarian whereby we associate numerals to our personal sensations of comfort/discomfort. Thirty degrees F is experienced as cold and 70 degrees F is considered warm. In countries using Celsius, 0 C and 20 C convey approximately the same sensations. The two scales, C and F, illustrated in the figure are not different. The distance between the two horizontal lines indicating high (red line) and low (blue line) show an equal vertical distance of length on the F and C scales. Any other lines drawn horizontally across the two tubes will indicate exactly the same elevation on both scales.

One intriguing aspect of this instrument is that volume in three dimensions for the thermometer has been reduced to length in one dimension for interpreting temperature. A complex variable has been reduced to a simple one. Rasch (1980) in discussing models in classical physics remarked,

> None the less it should not be overlooked that the laws do not at all give an accurate picture of nature. They are **simplified descriptions of a very complicated reality**" (p. 10, our emphasis).

This point seems rarely appreciated to judge from the voluminous amount of commentary in the social sciences citing how "complicated" reality is, and how difficult it is to model. Physics has progressed admirably following "simple" laws to model complex matters. Scientists appreciate complexity, but nature cannot be understood when complexity is made a stumbling block to understanding. In such instances, emphasizing complexity obfuscates understanding and knowledge. This temperature example reminds us that

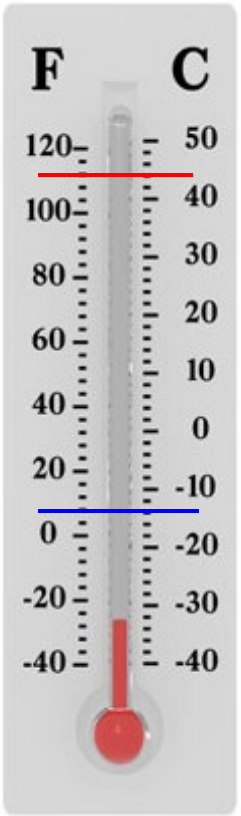


Figure 2. Celsius and Fahrenheit Temperature. From: www.scimathmn.org

complexity can be modeled in a simple fashion if only we can find a useful way to do so.

What is different between these two temperature scales is their *division* of length into segments, each one with different units and different origins for locating zero degrees. Celsius and Fahrenheit report different temperature numerals, but not different temperatures. It is the numerals that differ, not the temperature because the values of C and F can be connected by an algebraic expression, e.g. 9C = 5F – 160. Entering C and solving for F, or vice versa, give us the corresponding value.

A horizontal line across the picture of the temperature tubes supplies all the visual analogy we need to move from one scale to the other. This is because the "height" i.e. the length of mercury in the tube is invariant. It is the same height for each scale. The C and F scales are shown to be equal by observing this line connecting the two lengths. Algebra connects these two different scales precisely. What is not the same are the respective scale divisions and there are numerous variants.

The implications of this simple example can be important for understanding the essence of measuring:

1. We measure by analogy. We have moving hands, clocks ticking, and sand trickling through an hour-glass. No matter how sophisticated the device (cesium clock) analogy prevails in some form. *For temperature:* The internal liquid of a glass thermometer is a visual representation on the quantitative scale(s).

2. We should not be confused by differing scale values and origins into thinking complexity abounds. A validly constructed instrument emanates from a single, unified variable. The problem is to devise and construct one. *For temperature:* There is only one construct variable, but many ways to divide and express it.

3. Validity rests on achieving instrument integrity and invariance. Everything else is peripheral to this problem, and only serves to confuse the matter. Constructing the instrument and applying it to life are two entirely different matters not to be confused. *For temperature:* The instrument is foundational, applications follow.

4. Portability is necessary. Handled properly the instrument is useful in almost all locations. Extreme conditions in temperature or elevation above/below sea level require modifications and corresponding interpretation. *For temperature:* General application and utility constitute validity with some unique exceptions.

5. Utility is an important aspect of measuring. The choice between two explanations, complex vs. simple (Occam's razor), favors the simple as the useful one. Utility implies understanding. *For temperature:* Giving one's attention to observing the temperature, and not to the instrument illustrates the successful achievement of utility.

Mark Stone and Jack Stenner

References:

Crosby, A.W. (1997) The Measure of Reality. Chapter 1: Pantometry Achieved. assets.cambridge.org

Kant, I. (1917). *Gesammelte Schriften*, Vol. 7. Berlin: Reimer. (Original work published in 1798)

Oppenheimer, R. (1956). Analogy in science. *American Psychologist*, 127-135.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 4,* 321-333.

Rasch, G. (1980). *Probabilistic models for some intelligence and achievement tests*. Chicago: The University of Chicago. (Originally published 1960)

Stone, M., Wright, B. & Stenner, J. (1999). Mapping variables. *Journal of Outcome Measurement, 3 (4),* 306-320.

Vaihinger, H. (1935). *The philosophy of 'As If'*. London: Kegan Paul.

Georg Rasch and Item Fit

Georg Rasch didn't expect all kinds of items to fit a Rasch model in all kinds of frames of reference, so he was always extremely careful about testing the items to see whether there was something wrong with the way they had been constructed or with the theory underlying the items. He was in fact an absolute fundamentalist when it came to model checking not only for the Rasch model, but for all kinds of statistical models.

Graphical techniques were very important to Rasch. He had a principle saying (in Danish) that you should *"tegne"* before you *"regne"* (meaning "plots" before "calculations") and he had a lot of students (including Peter Allerup) doing the plots for him. He would never draw the ICC curves as we do today. Rasch plotted the logistic values of the probabilities against estimates of person parameters (or similar but more complicated functions of the item parameters against the total scores) because it is much easier to assess systematic departures from straight lines than departures from logistic curves (See Figure from Rasch, 1960).

In addition to these plots he would, of course, also use numerical tests and he would always insist that these calculations should be made relative to the conditional distribution of item responses given the total score to make sure that he had separated his inference on items from the persons.

You can find some of this in his 1960 book, but far from all. We know that he at some point worked on a sequel to the book that he never finished. When we celebrated his centenary in 2001 we published a collection of his unpublished papers and notes. (You can find this collection at www.rasch.org/memos.htm). Among these papers we included a chapter on "Estimation of Parameters and Control of the Model for Two response Categories" where he describes five different methods including a test for the hypothesis that the item discrimination is the same for all items and including much of the theory of conditional inference that Erling B. Andersen worked on and published during the 70'es. It is quite interesting reading so take a look at it. You can find it at www.rasch.org/memo196y.pdf

Rasch's view on item fit analyses were that evidence against the item means that it should be either revised (if at all possible) or removed. That goes both for items where item discrimination is too weak and items where discrimination is too strong (Infits and/or Outfits that are smaller than 1). That is also my point of view, but the interpretation of the lack of fit of the item is very different for items that do not discriminate and items with too strong discrimination. In the first case I would suspect inept item writing or multidimensionality. In the second case I would always look for evidence of local response dependence (LD) because I know that positive local dependence has the effect that the item discrimination of the items look stronger than expected by the Rasch model.

It is my experience this is the case in many of the analyses where I find evidence of too strong discrimination. You can find one such example in

Kreiner S (2011) A Note on Item-Restscore Association in Rasch models. Applied Psychological Measurement, 35, 557-561

where the local dependence is a consequence of inept item writing in the sense that items are phrased in such a way that local dependence is unavoidable.

Svend Kreiner
University of Copenhagen
Denmark

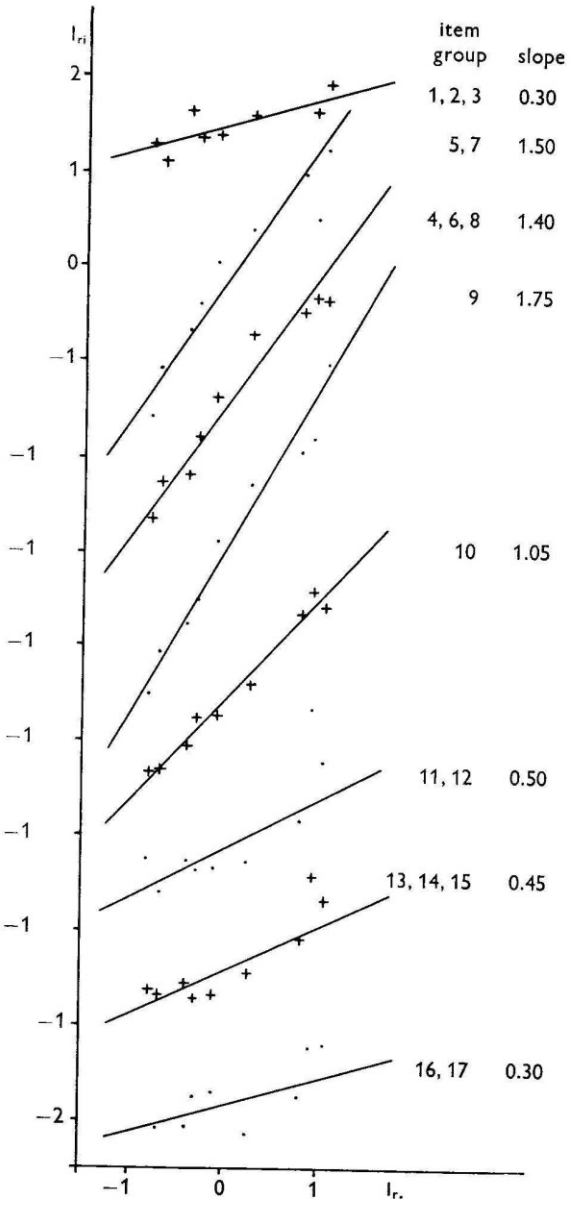


Figure 14. Subtest F of BPP. l_{ri} plotted against l_r for each group of items. From Rasch (1960) "Probabilistic Models for Some Intelligence and Attainment Tests".

2012 World Standards Day Paper Competition

Enter the Competition! Hard copy paper submissions (no emails accepted) must be received, with **an official entry form, by midnight August 10, 2012**, by the SES Executive Director, 1950 Lafayette Road, Box 1, Portsmouth, NH 03801. Cash prizes of US$2,500, $1,000, and $500 are awarded to the top three papers. Further information and the official entry form are available at www.ses-standards.org/displaycommon.cfm?an=1&subarticlenbr=77

How the Competition relates to us: This year's theme is *"Standards Increase Efficiency,"* in recognition of the fact that standards increase efficiency and reduce waste, not only in measurement but in any process or outcome affected by or involving measurement. Standards embody state-of-the-art know-how and so remove the need for every organization in a field or industry to master the latest techniques themselves. Further, because they are public and because they are established via consensus processes involving all stakeholders in an industry, standards even out unwanted variation in measurement quality. Finally, standards increase efficiency by establishing a common framework for decision making, outcome evaluation, and quality improvement, because the inferences made from quantitative comparisons are coordinated, aligned, and harmonized toward shared purposes, with no need for painstakingly negotiating the details of qualitative comparisons based in ordinal scores or percentages.

In education, for instance, because there are neither universally accepted uniform units nor instruments traceable to them, measurement quality varies greatly across classrooms, schools, districts, states, and commercial testing agencies. Teachers, principals, researchers, administrators, and psychometricians create tests and assessments individually and in groups, with massive amounts of duplicated effort, inefficiency, and differences in quality. Lacking uniform units and metrologically traceable instruments, educators are stuck, mired in a swamp from which it is impossible to even approach fulfilling their potential for developing innovative products and services.

The question is, if all educational measures were linked to common reference standards, what kinds of practical guidance could be provided on issues that would assist schools, districts, states, and curriculum developers in increasing their efficiency and effectiveness to meet the needs of students, teachers, parents, researchers, and employers in the coming years?

In the wider world, companies compete globally more effectively and efficiently, at lesser cost and with less waste, when they have consensus standard measuring units to inform their processes. The same is already true in many different ways in education, from the standards used in constructing school buildings and supplying their electricity, information, technology, and water, to the accounting standards used in budgets and purchasing, to the food quality and quantity standards informing cafeteria operations. The state-of-the-art know-how contained in standards is accessible to all, helps avoid duplication, and allows us to invest more in other priorities.

Given the proven state of the art in measurement theory and practice, and given the dire need for increased efficiency and reduced waste in education, health care, and social services, it's way past time uniform measurement standards were developed and implemented in these areas.

The 2012 World Standards Day paper competition is an opportunity for Rasch researchers to tell the story of how better measurement could impact the larger economic and social spheres of life. The standards community is interested in learning more, following on recent white papers published by NIST and NSF, and on last year's third place award to William Fisher for his paper, *"What the World Needs Now: A Bold Plan for New Standards,"* which will be the cover story in the forthcoming May/June issue of *Standards Engineering The Journal of SES—The Society for Standards Professionals*. A PDF of the paper is available at www.ses-standards.org/displaycommon.cfm?an=1&subarticlenbr=56.

William P. Fisher, Jr.

Mapping Differential Item Functioning (DIF Maps)

Variable maps provide useful tools for communicating the meaning of constructs in the human sciences. It has not been recognized that differential item functioning (DIF) can also be represented in a meaningful way on a variable map. In this case, the underlying continuum represents the differences between subgroups with comparable levels of achievement across a set of test items.

Data from Engelhard, Wind, Kobrin, and Chajewski (2012) are used to illustrate the concept of a DIF map. DIF was calculated as the difference in logits between

Figure 1. DIF Map for Gender

Males Females

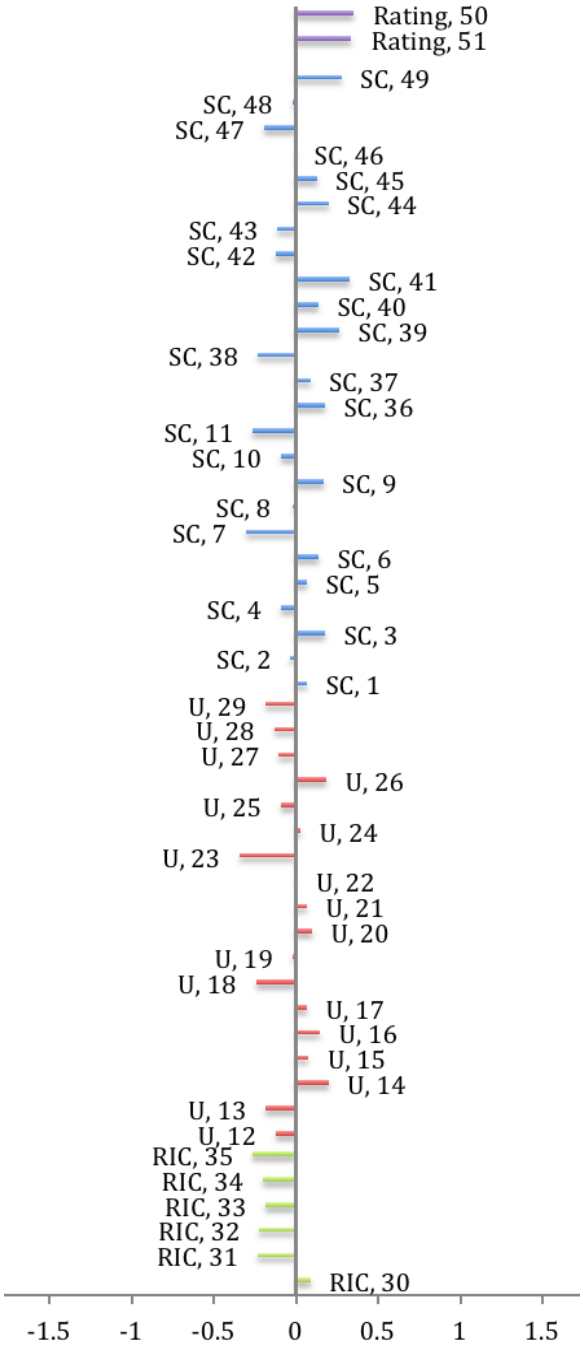
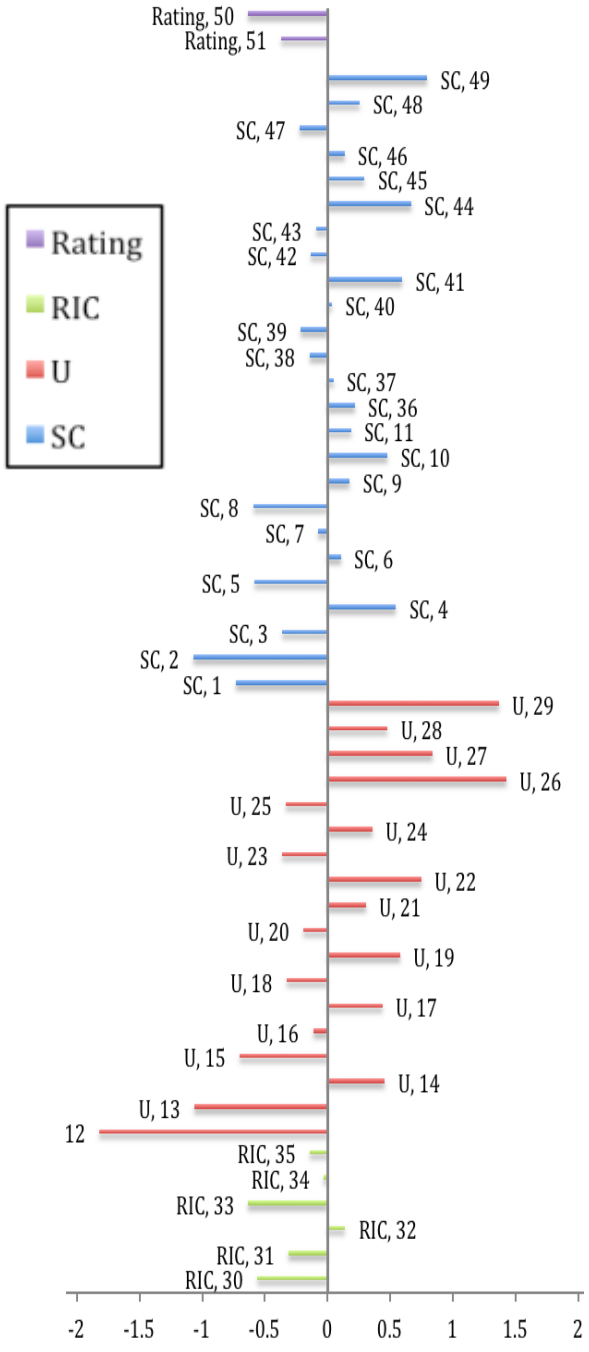


Figure 2. DIF Map for Best Language

English Best Language Another Language



Item Subsets: SC: Sentence Correction· U: Usage ·

RIC: Revision in Context· Ratings: Two Essay Ratings

separate item calibrations within subgroups based on the Rasch model. Two DIF maps are shown in Figures 1 (gender) and 2 (best language). The horizontal bars reflect the magnitude and direction of the differences between item calibrations for the comparison groups. The subset classification and item ID number for each SAT-W item are indicated on the DIF maps (SC=Sentence Correction, U=Usage, RIC=Revision in Context, and Rating= two separate ratings for the essay). There are several rules of thumb that can be used for interpreting the substantive significant of DIF, such as the half-logit rule proposed by Draba (1977). However, the reader is reminded that DIF maps stress the idea that DIF is a continuous variable, and that arbitrary cut points may not go far enough in aiding the substantive interpretation of DIF.

Figure 1 illustrates DIF in terms of gender subgroups. As can be seen in this figure, DIF appears to vary across item subsets, although the magnitudes of the gender differences are generally small. None of the items exhibit gender DIF based on the half-logit rule. Data were also collected on whether or not English was reported by the students as their best language. The magnitude and directionality of DIF are shown in Figure 2, and they are somewhat different from the DIF patterns shown in Figure 1. Since the SAT-W is designed to measure academic English, it is not surprising that several items exhibit DIF related to best language. For example, the English Best Language group has higher scores on both essay ratings as would be expected given the purpose of the assessment.

DIF analyses have become a routine part of the test development process (Zumbo, 2007). A variety of methods have been proposed for conducting DIF analyses, and all of the methods yield continuous indicators that can be used to create DIF maps. Rasch-based approaches (Wright, Mead, & Draba, 1976) are used here to guide the creation of the DIF maps.

[Acknowledgement: The College Board provided support for this research. Researchers are encouraged to freely express their professional judgments. Therefore, points of view or opinions stated in College Board supported research do not necessarily represent official College Board position or policy.]

Stefanie A. Wind and George Engelhard, Jr.
Emory University

Rasch Measurement Transactions
www.rasch.org/rmt
Editor: Kenneth Royal
Guest Editor: John Michael Linacre

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

References:

Draba, R. E. (1977). The identification and interpretation of item bias. (Research Memorandum No. 25). Chicago: Statistical Laboratory, Department of Education, University of Chicago.

Engelhard, G., Kobrin, J., Wind, S.A., & Chajewski, M. (2012). Differential item and person functioning in large-scale writing assessments within the context of the SAT Reasoning Test. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, CA.

Wright, B. D., Mead, R., & Draba, R. (1976). Detecting and correcting test item bias with a logistic response model. (Research Memorandum No. 22). Chicago: University of Chicago, MESA Psychometric Laboratory.

Zumbo, B.D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. Language Assessment Quarterly, 4(2), 223-233.

ORVOMS: Ohio River Valley Objective Measurement Seminar

The 2nd Annual Ohio River Valley Objective Measurement Seminar (ORVOMS) was held on May 18th, 2012 at the University of Kentucky. It was hosted by Dr. Kelly Bradley of the Department of Educational Policy Studies & Evaluation and co-sponsored by Dr. Arne Bathke of the Applied Statistics Laboratory. The keynote speaker was Dr. Richard Smith from Data Recognition Corporation. In addition to the regional attendees some participants traveled from as far away as Florida, Michigan, and Minnesota.

Presentations included:

Using Rasch measurement to inform policy and practice through comparisons of theoretical and empirical hierarchies

Equivalence of Angoff Ratings and Calibrations

An External Validation Study of a Classification of Mixed Connective Tissue Disease and Systemic Lupus Erythematosus Patients

Using the Rasch Rating Scale Model to Measure Job Satisfaction among Kentucky Head Principals

An Exploration of Data Driven Decision Making Among College Admission Professionals

When does Guessing Begin?

We thank everyone who participated and presented this year—it was an interesting and collegial meeting. We look forward to next year's seminar and hope that you will be able to participate. For information about upcoming events or to be placed on our mailing list please contact:

Melanie Lybarger
Psychometric Research Associate
The American Board of Family Medic**ine**
mlybarger ~ theabfm.org

Leniency of Raters across Time-Points

Hung and Wang (2012) report that 238 workers from were assessed on four occasions by five managers according to five job criteria (thoroughness, creativity, complexity, structure, and accuracy) along a 5-point rating scale.

WinBUGS was used to model the changes in rater leniency/severity. In the Paper's Figure 2, reproduced here, we see that the raters did not follow a predictable pattern across time. Always the most severe was Rater 2. If we assume that the true distribution of the workers are the same in each Department, then workers rated by Rater 2 are disadvantaged. Their ratings would always be lower than workers in the other Departments. This Paper supports the proposition that performance ratings must be adjusted for the severity of the raters.

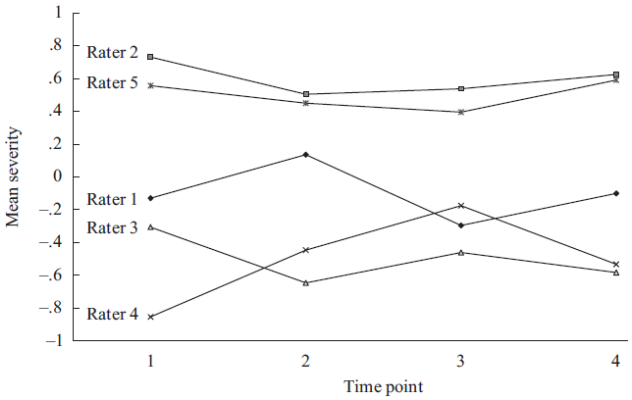


Figure 2. Mean severity for the five raters across time points in Hung & Wang (2012).

Hung, L-F, Wang, W-C (2012). The Generalized Multilevel Facets Model for Longitudinal Data. Journal of Educational and Behavioral Statistics, 37, 2, 231–255.

What the World Needs Now ....

"(Crosby, 1997, p. x) shows that the unity of mathematics and measurement in a quantitative model of the natural world explains why, between 1250 and 1600, Europeans "were able to organize large collections of people and capital and to exploit physical reality for useful knowledge and for power more efficiently than any other people of the time." It can be reasonably expected that the similar unification of mathematics and measurement in a quantitative model of the psychosocial world also will enable new magnitudes of efficiency and effectiveness to be achieved in caring human relations."

Fisher, W. P., Jr. (2012). What the world needs now: A bold plan for new standards. *Standards Engineering, 64, 3,* 1-5.

Reference: Crosby, A. W. (1997). *The measure of reality: Quantification and Western society, 1250–1600.* Cambridge: Cambridge University Press.

A Comment on the H^T Person Fit Statistic

The non-parametric H^T fit statistic (Sijtsma, 1986) is a transposed formulation of Loevinger's H scalability coefficient for items. It is evaluated by Karabatsos (2003). He reports "Overall, the H^T statistic is best [of 36 person fit statistics] in identifying aberrant test respondents. It is also among the best in detecting each of the five different types of aberrant-responding examinees, and in detecting such examinees in each of the short, medium, and long test conditions."

H^T is defined for the rows (persons) of a complete rectangular dichotomous dataset.. Let us focus on person *n* in a dataset which has L items and N persons. Then, following Karabatsos (2003),

$$H^T(n) = \frac{\sum_{m=1, m\neq n}^{N} \left(\left[\sum_{i=1}^{L} X_{ni} X_{mi} \right] / L - P_n P_m \right)}{\sum_{m=1, m\neq n}^{N} \left(\min\left[P_n (1 - P_m), P_m (1 - P_n) \right] \right)}$$

where X_{ni} is the scored (0,1) response of person *n* to item *i,* and $P_n = S_n/L$ where S_n is the raw score of person *n*, and similarly for person *m.*

H^T is the sum of the covariances between person *n* and the other persons divided by the maximum possible sum of those covariances, so that the range of H^T is –1 to +1. When the responses by person *n* are positively correlated with all the other persons, then $H^T(n)$ will be positive. When person *n* is negatively correlated with all the other persons, then $H^T(n)$ will be negative. When person *n*'s responses are random, $H^T(n)$ will be close to zero. When the data fit the Rasch model, we expect $H^T(n)$ to be somewhat positive, because all the person response strings will correlate positively with the item-easiness hierarchy, and so positively with each other.

According to Karabatsos (2003), "the critical values $H^T \leq$ .22 best identify aberrant-responding examinees." In my own informal analyses, the correlation between H^T and the Rasch Infit mean-square was about –0.9, but I was unable to identify a unique Infit mean-square value corresponding to H^T = .22. Since Sijtsma and Molenaar (2002) report 0.3 to be a critical value for coefficient H, a small simulation study may be required to determine the critical value of H^T for a specific empirical dataset.

John Michael Linacre

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. Applied Measurement in Education, 16(4), 277-298.

Sijtsma, K. (1986). A coefficient of deviant response patterns. Kwantitative Methoden, 7, 131–145.

Sijtsma K,, Molenaar I.W. (2002). Introduction to Nonparametric Item Response Theory. Thousand Oaks, CA: Sage.

6th Annual UK Rasch Users Group Meeting

20th March 2012, Weetwood Hall, Leeds, UK

Svend Kreiner. University of Copenhagen. Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment.

Christine Merrell , Peter Tymms and Irene Styles. Universities of Durham, and Western Australia. Cross-country Comparisons of Inattentive, Hyperactive and Impulsive Behaviour in School-Based Samples of Young Children.

Maria Pampaka, The University of Manchester. Measuring Pedagogies in Mathematics with the Rasch Model: from Secondary School to University and across countries.

Fabio Camargo, Brian Henson. University of Leeds. Rasch Theory in Product Design Applications

Bryan. Moreton. University of Nottingham. Rasch Analysis of the Intermittent and Constant Osteoarthritis Pain Questionnaire.

Ashworth, F., Bauch, E., Bateman, A. Oliver Zangwill Centre for Neuropsychological Rehabilitation. Forms of Self-Criticism, Self-attacking and Self Reassurance Scale in an ABI population using Rasch Analysis

Tracey Young, John Brazier, Donna Rowen, Brendan Mulhern, Ifigeneia Mavranezouli. ScHARR, University of Sheffield and University College London. Deriving preference-based utility measures from existing measures: How Health Economists Make Use of Rasch Analysis

Bateman, A. Sun, L . Oliver Zangwill Centre for Neuropsychological Rehabilitation and University of Cambridge. Analysis of Responses to the Ekman 60 Faces: Perception of Emotion Testing in 194 patients who have suffered from brain injury.

Panayiotis Panayides . Lyceum of Polemidia. Is Aberrant Response Behaviour an Inherent Characteristic of Students Taking Classroom Maths Tests?

Twiss J, Crawford SR, McKenna SP. Galen Research, Manchester. Co-Calibrating Scores from Two Dermatology-Specific Patient Reported Outcome Measures.

UK Rasch User Group, www.rasch.org.uk

In the realm of scientific creativity ...

"Quality is a probabilistic function of quantity."

Simonton, D. (2003), Scientific Creativity as Constrained Stochastic Behavior: The Integration of Product, Person and Process Perspectives. Psychological Bulletin, 129(4), 475-494.

Report on Ben Wright

As of May 29, 2012, Benjamin D. Wright remains at the Warren Barr Pavilion, 66 West Oak Street, Room 615, Chicago, IL 60610. (312) 705-5100. His room is in the "Avalon Wing," which has a very good staff. I expect this to be his last home, unless Andy Wright or one of Ben's other children, Amy, Andy, Chris and Sara, take him to New York. That seems unrealistic. Such a trip would be hard on him.

He continues to do well when visitors come. Although, more and more he sleeps. Recently, a couple of times he did not wake up when I was there. I have been seeing him less than I did but still make it once a week or so. A family member visits from NY, usually once a month. Not a lot of former students come by but I happened to meet Filemon Cerda when he came by about a two weeks ago.

I am making progress on Ben's biography and should have a preliminary draft ready to send to an editor very soon. It is hardly comprehensive. I have more material but this draft will mainly address ancestral and early childhood influences that played out in his choice of career and in the contributions he made. Not sure how others will respond but Ben really likes it when I read sections to him. A few years before his cerebral incident in 2001, Ben had put a lot of effort into researching his family tree and their activities.

Ed Bouchard, co-author with Ben Wright of "Kinesthetic Ventures" www.uprighting.com/introduction.pdf

Journal of Applied Measurement
Vol. 13, No. 1, 2012

Formulating Latent Growth Using an Explanatory Item Response Model Approach. *Mark Wilson, Xiaohui Zheng, and Leah McGuire. 1-22*

Using the Mixed Rasch Model to Analyze Data from the Beliefs and Attitudes About Memory Survey. *Everett V. Smith, Jr., Yuping Ying, and Scott W. Brown, 23-40*

An Examination of Personality Characteristics Related to Acquiescence. *Christine DiStefano, Grant B. Morgan, and Robert W. Motl, 41-56*

Construction and Validation of Two Parent-Report Scales for the Evaluation of Early Intervention Programs. *William P. Fisher, Jr., Batya Elbaum, and W. Alan Coulter, 57-76*

Multi-Factor Scale Consolidation When Theory is Weak. *Nikolaus Bezruczko and Kyle Perkins, 77-96*

Understanding Rasch Measurement: Developing an Emotional Distress Item Bank for Cancer Patients. *Allen W. Heinemann, Rita K. Bode, Sarah Rosenbloom, and David Cella, 97-113.*

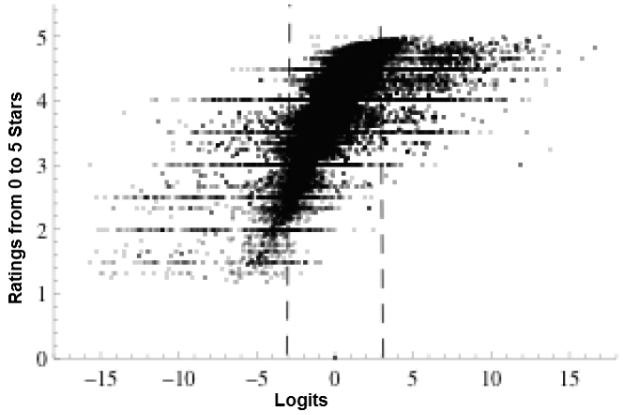
Richard M. Smith, Editor, www.jampress.org

Rasch and Distributions

Question: What assumptions are made about distributions by the Rasch model?

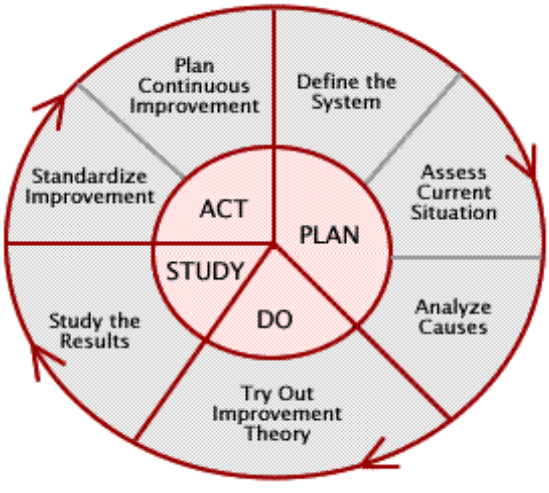
Answer: The Rasch model makes no assumptions about the distributions of the parameters. Maximum Likelihood Estimation (MLE) assumes that the randomness in the data is normally distributed. Some estimation methods, such as Marginal Maximum Likelihood (MMLE), assume that the incidental parameters (usually persons) conform to a well-behaved distribution (usually normal).

Adjusting for Rater Leniency



"Figure 3, which plots the average number of stars awarded (Y-axis) as function of the Rasch product parameters (X-axis). It can be seen that the relation between these two variables is decidedly non-linear because raw differences near the top underestimate the true differences, thereby again calling into question the use of raw scores (presently, the number of stars) as quantitative indices."

Lange R., Lange X. (2012) Quality Control in Crowdsourcing. AAAI Spring Symposium Series, 2012



W.E. Deming's Improvement Cycle
University of Illinois at Chicago
Division of Specialized Care for Children..
www.uic.edu/hsc/dscc

Rasch-related Coming Events

July-Nov., 2012 On-line course: Introduction to Rasch Measurement of Modern Test Theory (D. Andrich, RUMM2030), Perth, Australia, www.education.uwa.edu.au/ppl/courses/introduction

July 6 - Aug. 4, 2012, Fri.-Sun. On-line course: Practical Rasch Measurement - Further Topics (E.V. Smith, Winsteps), www.statistics.com/raschfurther

July 10-12, 2012, Tues.-Thurs. Psychometric Society Annual Meeting, Lincoln, Nebraska, www.psychometrika.org

July 15, 2012, Sun. Online Degree Programs: Application deadline: Measurement, Evaluation, Statistics and Assessment (E.V. Smith), MESA, University of Illinois at Chicago www.uic.edu/gcat/EDMESA.shtml

July 16, 2012, Mon. Online Degree Programs: Application deadline: Rasch Measurement of Modern Test Theory (D. Andrich), Pearson Psychometrics Laboratory, University of Western Australia. www.education.uwa.edu.au/ppl/courses

July 23, 2012, Mon. Submission deadline: AERA Annual Meeting, San Francisco, www.aera.net

Aug. 4-5, 2012, Sat.-Sun. PROMS2012 Workshops (Jim Sick, Eric Wu, Trevor Bond, Jack Stenner), Jiaxing University, Zhejiang Province, P.R.China,

Aug. 6-9, 2012, Mon.-Thur. PROMS2012, Jiaxing University, Zhejiang Province, P.R.China, cfs.zjxu.edu.cn/proms

Aug. 10 - Sept. 9, 2012, Fri.-Sun. On-line course: Many-Facet Rasch Measurement (E.V. Smith, Facets), www.statistics.com/facets

Aug. 12-14. 2012, Sun.-Tues. IACAT 2012, International Association for Computer Adaptive Testing, Sydney, Australia, www.iacat.org

Aug. 13-17, 2012, Mon.-Fri. On-line short course: Applied Measurement with jMetrik (P. Mayer), curry.virginia.edu/community-programs/conferences/jMetrik

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

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,

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, www.leeds.ac.uk/medicine/rehabmed/psychometric