RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 26 No. 4 Spring 2013 ISSN 1051-0796

DIF Sample Size Nomogram

Differential Item Functioning (DIF) compares the difficulty of an item for a person sample of interest, the Focal Group (F), with the difficulty of the item for another group, the Reference Group (R). As we plan to investigate DIF, several questions arise:

1) What difference between the two item difficulties, i.e., what DIF size, is large enough to have substantive consequences? If we lack better information about the situation, we may choose the DIF size specified by Educational Testing Service which is 1 Delta δ unit = 0.426 logits (see RMT 20:3 p. 1070), rounded up to 0.5 logits for sample-size purposes.

2) What difference between item difficulties is unlikely to be merely a chance result, i.e., when is the DIF size statistically significant? We usually choose the conventional statistical significance level, $p \leq .05$ (Fisher, 1926).

3) How many members of the Focal and Reference Groups are required for the DIF size specified in (1) to meet the DIF significance criterion specified in (2)? What samples sizes are required? This nomogram provides a guide. The nomogram is based on a Student's *t*-statistic with two independent groups of different sizes, N_F and N_R, but equal standard deviations, S. The *t*-statistic of the DIF size, D, is

$$t = \frac{D}{S}\sqrt{\frac{N_F N_R}{N_F + N_R}}$$

Group size values are chosen such that the two-sided probability of observing *t* or greater is $p = .05$ for the specified DIF size, D. The standard deviation, S, is chosen at a reasonable value, 1.88 logits, based on inspection of empirical data.

According to the nomogram, confirming a DIF size of 0.25 logits between Reference and Focal groups requires the groups to total at least 2 x 1,500 = 3,000 persons. 0.5 logits DIF size requires at least 2 x 300 = 600 persons. 1.0 logits DIF size requires at least 2 x 100 = 200 persons.

John Michael Linacre

Fisher R.A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture, 33*: 504.

Tristan, A. (2006). An adjustment for sample size in DIF analysis. *Rasch Measurement Transactions, 20*:3, p. 1070-71.

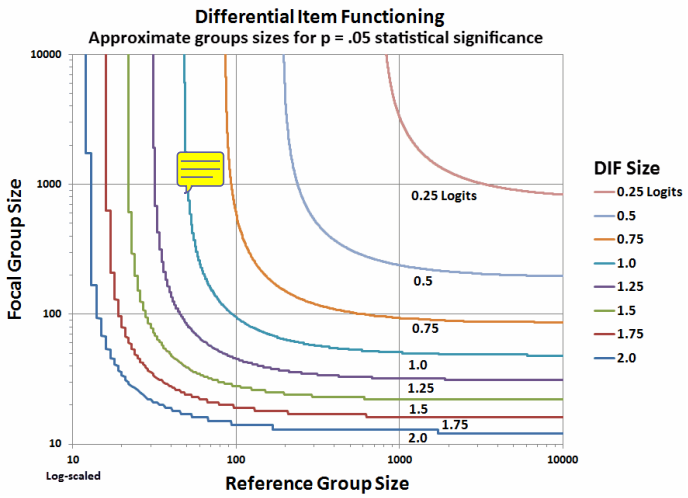


Table of Contents

Hanning Person Response Functions

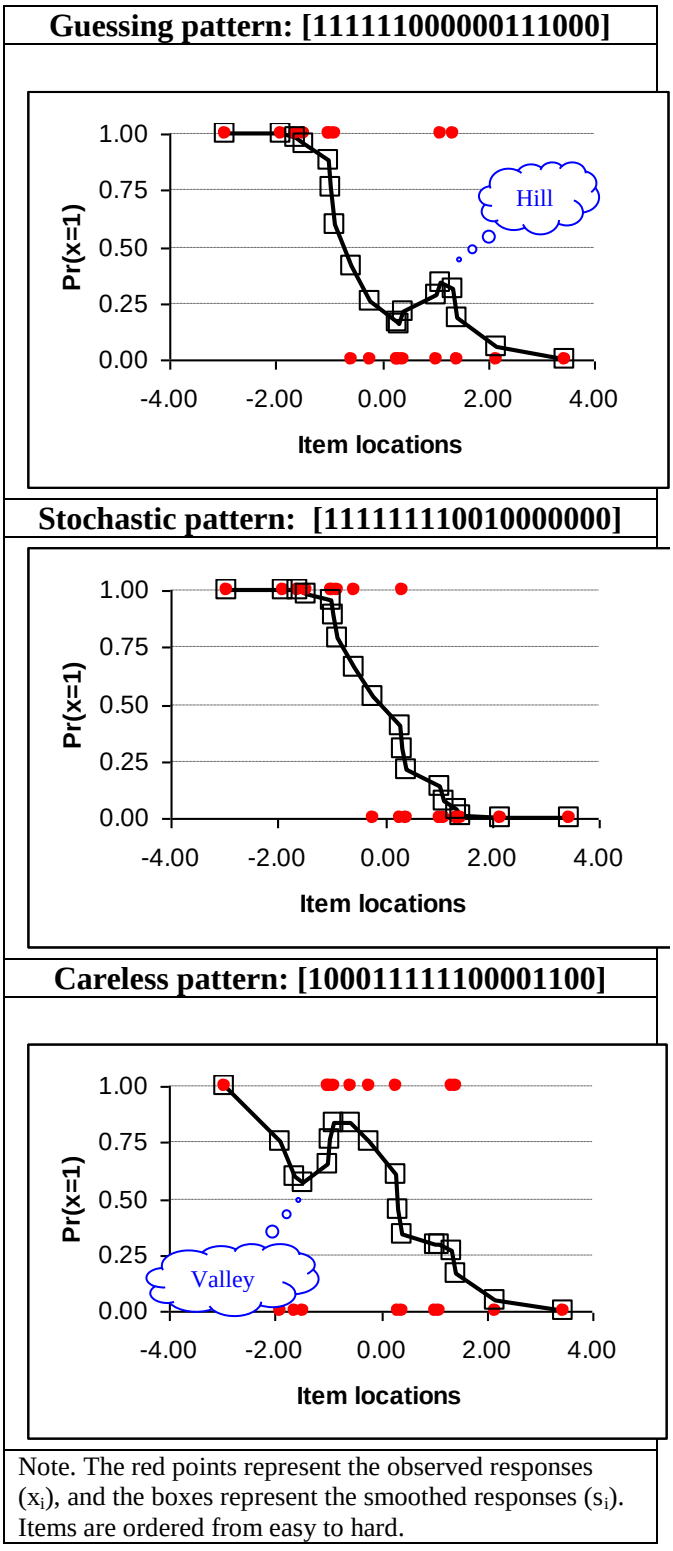
Person response functions (PRFs) offer an approach for providing additional information when aberrant person response patterns are encountered. PRFs can be based on parametric IRT models, such as the Rasch model. Nonparametric PRFs offer another way to estimate PRFs that may reveal unusual aspects of a person's responses to a set of items including non-monotonic relationships between item locations and person responses. A promising estimator for smoothing functions called "hanning" has been suggested by Tukey (1977) based on a method proposed by von Hann (1903). This estimator provides a nonparametric approach for modeling a functional relationship between calibrated item difficulties and person responses. A simple version of this algorithm recommended by Velleman and Hoaglin (1981) is given by

$$s_i = (y_{i-1} + 2y_i + y_{i+1}) / 4 \quad [1]$$

where y_i is replaced by s_i. In the context of person response functions, the sequence of values that define the x-axis are based on item calibrations, and the values on the y-axis to be smoothed by Equation 1 are the dichotomous person responses (where y_i = 0,1 with 0 representing an incorrect response and 1 representing a correct response to item *i* by a person). These smoothed values offer a promising approach for illuminating various aberrant and unexpected responses through graphical displays. In essence, the first iteration reflects empirical proportions (weighted) of number correct responses for subsets of four items. The smooth values can continue to be smoothed in an iterative fashion.

In order to illustrate the hanning of person response patterns, item calibrations from Engelhard et al. (in press) are used. The usage items (N=18) for assessing writing from the 2009 administration of the SAT writing assessment define the latent variable (x-axis). Nine iterations of Equation 1 are used to yield the smoothed values in this study.

Figure 1 presents three possible response patterns for students with average scores (50% correct): a guessing pattern, a stochastic pattern, and a careless pattern. The guessing pattern shows a distinctive "hill" reflecting unexpected success on harder items. The careless pattern shows a "valley" with unexpected failure on easier items. The stochastic pattern shows responses that would exhibit good fit to a parametric PRF. PRFs can display hills and valleys in response patterns that may provide the basis for proposing substantive hypotheses, such as guessing and careless behaviors, that merit future study.



Note. The red points represent the observed responses (x_i), and the boxes represent the smoothed responses (s_i). Items are ordered from easy to hard.

Figure 1: Person response functions

In summary, person response functions provide graphical displays that convey information about idiographic aspects of each person's response pattern. Standard errors of measurement for estimated person scores are not sufficient for reflecting the uncertainty in each person's

response pattern. Person fit indices are helpful, but multiple response patterns can yield comparable misfit values for most person fit statistics. Person fit statistics do not provide unequivocal interpretations of response patterns without additional inquiry. As pointed out by Tukey (1977),

> a basic problem about any body of data is to make it more easily and effectively handleable … anything that makes a simple description possible makes the description more easily handleable [and] anything that looks below the previously described surface make the description more effective (p. v)

Hanning PRFs is an approach for making person scores more "handleable" by displaying the potential functional relationships that may explain individual response patterns.

George Engelhard, Jr.
Emory University

References

Engelhard, G., Kobrin, J., Wind, S.A., & Chajewski, M. (in press). *Differential item and person functioning in large-scale writing assessments within the context of the SAT Reasoning Test.* College Board Research Report.

Tukey, J.W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley Publishing Company.

Velleman, P.F., & Hoaglin, D.C. (1981). *Applications, basics, and computing of exploratory data analysis.* Boston, MA: Duxbury Press.

von Hann, J. (1903). *Handbook of climatology.* New York: The Macmillan Company.

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

A Note from the Rasch SIG Chair

Greetings Rasch SIG colleagues,

I want to take this opportunity to provide a brief update on SIG membership and upcoming activities. This is an exciting time of year as we tiptoe closer to the annual AERA conference and SIG Business Meeting. Over the course of the past year I'm happy to report that the Rasch community continues to thrive and provide sound measurement solutions across an exceptionally broad range of applications. I believe this ongoing enthusiasm continues to be firmly exemplified through this SIG, *RMT*, *JAM*, Rasch Listservs, and other publications and professional organizations dedicated wholly, or in part, to this pursuit.

First, I want to acknowledge the ongoing efforts of Daeryong Seo and Kelly Bradley for their work as the Rasch SIG Program Co-Chairs for this year's meeting. I'm informed that this year's program looks fantastic and we can look forward to two paper presentations and a round table based on 19 accepted paper submissions. Secondly, I'm very happy to report that past Rasch SIG Chair Ed Wolfe (2008-2009) has agreed to speak at this year's business meeting. The focus of Ed's presentation will be "Bringing Together Cognitive and Psychometric Models of Rater Effects." As for this year's business meeting, you can expect it to be in line with the past few in that I will be providing a brief State of the SIG address prior to introducing Ed. The meeting is scheduled for Saturday, April 27th from 6:15 to 7:45. Hors d'oeuvres and a cash bar will be provided. I will send out more detailed information on all presentations and logistics prior to the AERA conference.

Third, I want to draw attention to a couple of the changes that have occurred over the past year and also which will be facing us in the near term future. *RMT*, as you know, has now existed as a foundation repository for ongoing technical discourse and evolution of Rasch measurement for decades. This past year saw the transfer of editorship from Mike Linacre to Ken Royal and I wanted to express my appreciation to both for ensuring this publication continues. While technically listed as a newsletter, the archives contain valuable technical contributions that are as applicable today as ever. Judging from the first couple editions under new editorship, I am excited to see how *RMT* not only stays true to the original intent but also offers a bit of new direction. Thanks indeed to both Mike and Ken!

Also, you have likely been well aware that all SIG bylaws have been going through revisions over the past year in order to conform to District of Columbia nonprofit laws and policy standards. Over the past several months these have been vetted through legal and are currently going through the final stages of revision and review. As of the

end of February I was told that we are very close to moving to completion. Stay tuned, as it will be our turn to vote on whether or not to accept the revised Rasch SIG bylaws. This vote is conducted by AERA, so expect to be contacted. The most significant revisions to our bylaws are that we will move to having three elected officers instead of two and that we will initiate an awards program recognizing the efforts of both early career and senior scholar accomplishments in Rasch measurement. Assuming the vote is favorable, the new bylaws would take effect immediately.

Lastly, I want to appeal to you all in terms of your continued membership and involvement. I firmly believe that the SIG is in excellent standing and in the midst of changes that will bring more opportunity, recognition, and reward to members. I've noted the addition of one more officer to our ranks in the next election cycle. For this year we have two needs to fill. One is for a volunteer to serve as our webmaster. If you are skilled in this area and willing to help out, please do reach out to us. This would entail occasional updates to the website and electronic posting of *RMT* articles. It is not a huge commitment, but one that would be very much appreciated. The second need is for a Program Co-Chair. Each year one of two Co-Chairs finishes his or her two year term with us. Co-Chairs coordinate the call for proposals for the annual meeting. If you would like to be considered for this role or have a nomination in mind, please contact any of the SIG officers.

Beyond this year's requests, there will certainly be additional opportunities to get directly involved in SIG activities down the road. On that note I will close with one final appeal. Membership is always somewhat dynamic and so far this year appears in line with the recent past. As I led off on an optimistic note, I want to parlay that and ask that you spread the word to others who may not be fully aware of SIG membership. Graduate students, junior professionals, lifelong scholars, interested practitioners are welcome and can each benefit through membership. Given this time of year when many are renewing annual memberships and choosing other associations, simply passing along a word of consideration can make a difference.

Thanks to you all and I look forward to seeing you in San Francisco this April!

Tim O'Neil
Rasch SIG Chair (2012-2014)

Should Woodcock's Test Design Nomograph be Adjusted and Applied to Polytomous Tests?

Woodcock's test design nomograph helps construct dichotomous tests (Woodcock, 1992). The computation of Woodcock's nomograph was reported by Pedler (1993) and helped us predict an estimated person logit standard error at a test's center based on item length, and logit range of a uniform test. This prompted us to use a simulation to examine whether Woodcock's expected SE values could be exactly applied to not only dichotomous tests, but also polytomous tests.

We found Woodcock's test design nomograph cannot be directly applied to both dichotomous and polytomous tests ($p < 0.001$) (see Table 1). Instead, the adjusted formula (regressing Woodcock SE value to predict Rasch minimal sample logit SE) can be feasibly and optimally useful while using item length with different types of category numbers and the logit ranges to predict a person logit SE at a test's center. Dichotomous scales are statistically significantly different from the polytomous with regards to the expected SEs. In contrast, there are no statistically significant difference between polytomous scales (Figure 1).

A polytomous item of m ordered categories contains m-1 dichotomous category boundaries (Linacre, 2000). A test of k items contains C_k - k dichotomous items. The number of active categories in the known test is shown as below:

$$C_k = \sum_{i=1}^{k} m_i$$

Table 1. Tests of Between-Subjects Effects

Source	Sum of Squares	DF	Mean Square	F	P
Woodcock's test design nomograph					
A:category	11.873	3	3.958	290.483	< 0.001
B:length	0.05	2	0.025	1.834	0.160
A*B	0.0941	6	0.016	1.151	0.330
Residual	42.168	3095	0.014		
Adjusted by regressing Woodcock to predict minimal sample SE§					
A:category	0.00003	3	0.00001	0.0251	0.995
B:length	0.003	2	0.00152	3.271	0.038
A*B	0.003	6	0.0005	1.086	0.368
Residual	1.435	3095	0.0005		

§ Regression equations for 4 category number (2 to 5) scales:

Y_Cat2 = 0.024 + 0.80 * Woodcock SE

Y_Cat3 = 0.007 + 0.547 * Woodcock SE

Y_Cat4 = 0.00969 + 0.4576 * Woodcock SE

Y_Cat5 = 0.00786 + 0.3984 * Woodcock SE

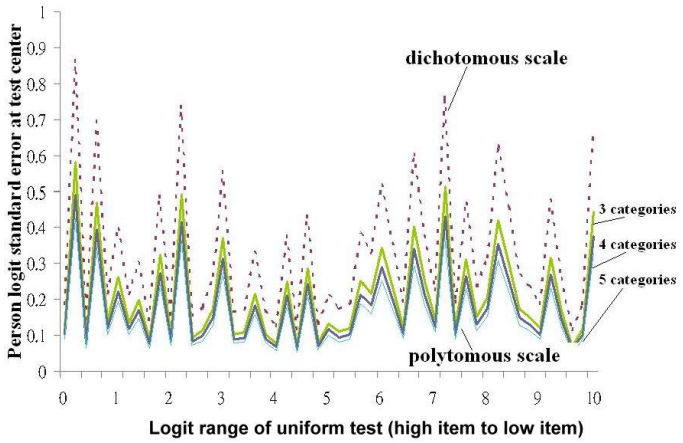


Figure 1. Comparison of Woodcock's test for SE by category number of a scale

Tsair-Wei Chien, Chi Mei Medical Center, Taiwan
Jianfang Zou, Academy of Medical Science, Shandong, China

References

Woodcock, R. W. (1992). Woodcock test construction nomograph. *Rasch Measurement Transactions, 6(*3), 243-244. Available at: http://www.rasch.org/rmt/rmt63p.htm

Pedler, P. (1993). Computerizing Woodcock's test construction nomograph. *Rasch Measurement Transactions*, 6(4), 255.
Available at: http://www.rasch.org/rmt/rmt64f.htm

Linacre, J.M. (2000). Predicting reliabilities and separations of different length tests. *Rasch Measurement Transactions*, 14(3), 767.
Available at: http://www.rasch.org/rmt/rmt143j.htm

Among Sticks and Axioms

Ben Wright sometimes used the analogy of sticks when discussing measures. *Best Test Design* (Wright & Stone, 1979) was to begin with such a discussion, but we changed to another analogy—the arrow. Later, *Measurement Essentials* (Wright & Stone, 1996) included a figure encompassed by sticks. The sticks have since come to serve as a *primitive measure* analogy useful to illustrate constructing variables. This *urgrund* of sticks becomes a foundation to making measures. The sequence is as follows:

1. Comparison is key; i.e. equal, greater, less.
2. Compare a chosen stick (agent) to an object. The stick is equal to, greater or less.
3. If greater than, choose a longer stick and compare again; equal to, greater, or less.
4. If less than, choose a shorter stick and compare; equal to, greater, or less.
5. Select as many sticks (agents) as required, longer or shorter, to "box-in" the object.
6. When the sticks chosen to use are selected and ordered with an equal difference between each of them, you have created a unit.
7. The more sticks chosen, the finer the difference (unit) by which to "box-in" the object.
8. Assemble the sticks in their graduated order. Comparison and order are shown to be essential to this process.
9. If one could collapse all the individual sticks into a single one; instead of an assemblage of sticks there would be a "ruler" constructed from the sticks.
10. Assigning numbers to this orderly sequence of sticks produces a "stick ruler" similar to a foot ruler demarcated by inches, or a yardstick by inches/feet, rod by feet, mile by feet, rods, etc.

The analogy of sticks provokes the construction of spelling items, math problems, indicants of depression, fear, anxiety, shame, etc. This approach is further illustrated and confirmed by observing the decreasing lengths of marimba key-length for unit scale differences in well-tempered pitch, likewise for varying string length observed on the harp or a concert grand piano, and the decreasing lengths of organ pipes. It is further manifested by experiments dating back at least to the Greeks, if not earlier, who used the monochord as the basis for the study of "music" (more correctly the fundamentals of today's acoustics) comprising one science in the medieval quadrivium.

Unique aspects for developing measures were employed much earlier according to archeological findings. One example comes from the Early High Period (c. 3200-2800 BCE) of the ancient Middle East documented by Nissen (1988) who identified a "unit bowl" used for dispensing the daily food ration across the entire Babylonian Empire. The proliferation of these unit bowls throughout the empire testify to a standard economic unit identifiable not just by its unit volume, but more so by its unique construction to designate its singular purpose, and keep it distinct from all other pottery. Furthermore, the pictograph of this bowl together with a head indicates "to eat."

An orderly arrangement of sticks can be associated with the axioms of quantity suggested by Höelder (1901) as given by Nagel (1931):

Nagel's axioms	Stick order as numbered above
1. Either a > b, or a < b, or a = b.	1, 2, 3, 4
2. If a >b, and b > c, then a > c.	5
3. For every a there is an a' such that a = a'.	2, 3, 4
4. If a > b, and b = b' , then a > b'.	2, 3, 4

5. If a = b, then b = a.	1, 2, 3, 4
6. For every a there is a b such that a > b (within limits).	2, 3, 4
7. For every a and b there is a c such that c = a + b.	8, 9
8. a + b > a'.	8, 9
9. a + b = a' + b'.	2, 3, 4
10. a + b = b + a.	1
11. (a + b) + c = a + (b + c).	8, 9
12. If a < b, there is a number n such that na > b (also within limits).	8, 9

These axioms can be somewhat aligned to the sequence of sticks as indicated in the column to the right of the axioms. The association need not correspond perfectly to substantiate the value of the "sticks" analogy in demonstrating how useful a primitive form of determining measures can be in producing a model for measuring. The alignment is not nearly as important as the fact that the sticks and axioms are both orderly constructions illustrating their importance in systematic fabrication. "All measurement rests upon having a qualitative ordering of the set of objects," write Luce and Narens (1981, p. 215). Order by comparison is essential to any construction.

We should not slight nor dismiss the role of simple strategies for conveying sophisticated processes. Once considered the domain of *analysis*, elementary school students are now exposed to previously advanced concepts such as the commutative, associative, distributive and transitive processes. Leopold Kronecker is said to have attributed creation of the integers to God, and all other mathematics to creation by man. We are a fabricating people.

Michell (2003, p. 300) provides some similar conditions which "characterise (sic) length as a continuous quantitative variable:

1. For every pair of lengths, a and b, one and only one of the following is true:
   (i) a = b;
   (ii) there exists another length c, such that a = b + c;
   (iii) there exists another length d, such that b = a + d.
2. For any lengths a and b, a + b > a.
3. For any lengths a and b, a = b = b + a.
4. For any lengths a, b, and c, a + (b + c) = (a + b) + c.
5. For any length a, there is another length, b, such that b < a.
6. For any lengths a and b there is another length, c such that c = a + b.
7. For every non-empty class of lengths having an upper bound, there is a least upper bound."

While these conditions suggest "length" in their specifications, the sticks "produce" length! The difference is considerable inasmuch as specification denotes a retrospective mathematical process by means of axioms, whereas the sticks fabricate length as a measure appearing before our eyes. The value of axiomatic specification is to provide a succinct, internally consistent process of logical steps, but Guttman (1971, p. 346) wrote, "Even in mathematics, axiomatization is an intermediate developmental stage; one must first have some idea of some body of inter-relationships for which the axiomatization may be fruitful." The ideals of exactness and rigor in mathematics are the product of time and refinement. The order of sticks manufactures length from operations that evolved early and developmentally; the same process by which any variable might be initially produced. Instrument refinement requires theory, continuous quality control, validation, etc., but that is another story.

A sense of cold and heat is crudely qualitative. We feel "hot" or "cold." Measures of temperature are made quantitative by fashioning a graduated tube constructed of glass together with mercury contained in a vacuum. This approach follows from numerous experiments using water, alcohol, etc. The instrumentation improved because the goal remained constant. Today it is a uniform association between the expansion of mercury and a measure of temperature. The sensations of cold and heat are derived from the human organism while temperature is manufactured analogously by an instrument -- a thermometer; a process little different from using sticks. We "sense" the temperature according to what thermometer scale we most commonly employ. Wittgenstein (1958, § 508) writes, "I am not used to measuring temperatures on the Fahrenheit scale. Hence, such a measure of temperature 'says' nothing to me." F-70 means pleasant and F-32 means freezing to those who are familiar with the scale. The matter rests upon an analogy, and the associations that are important for interpreting a measure. A NexTemp® (2004) thermometer is strikingly similar to the sticks analogy, but uses "chemical cavities" instead of sticks to "box in" one's temperature.

The sensation of cold and heat is experienced and crudely categorized. Temperature is constructed by employing comparison and order. A sensation usually lacks clear lines of demarcation further hampered by the "swamp of language." Thermometers employ a sequence of units (numbered) which, when correctly constructed and employed, produce an unambiguous result. A sensation seeks clarity as though residing in a fog, while a working thermometer produces a consistent, useful value. When Chang (2004) selected *Inventing Temperature* as the title of his book he proclaimed the essence of measuring. Measuring is invention, the process of variable construction, and that inventive process is never ending.

There is reality, and there is one's idealized goal. We might ponder G. H. Hardy's remark that "nothing practical" would occur in his *Course of Pure Mathematics* (1908), except "constructing" a world of mathematics.

We dwell in a fabricated land straddling two realms – the real and the ideal.

Mark Stone and Jack Stenner

References

Chang, H. (2004). *Inventing temperature*. New York: Oxford University Press.

Guttman, L. (1971). Measurement as structural theory. *Psychometrica, 36*(4), 329-347.

Hardy, G. H. (1908). *Course in pure mathematics*. London: Cambridge University Press.

Höelder, O. (1901) Die Axiome der Quantität und die Lehre von Mass. Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, *Mathematisch-Physische Klasse, 53*, 1-46.

Luce, H. D. & Narens, L. *Axiomatic measurement theory*. SIAM-AMS Proceedings, 13, 213-235.

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

Nagel, E. (1931). Measurement, *Erkenntnis*, Band 2, 313-333.

NexTemp® (2004). Medical Indicators, Inc. @medicalindicators.com.

Nissen, H. (1988). *Early history of the ancient near east*. Chicago: The University of Chicago Press.

Wittgenstein, L. (1958). *Philosophical investigations*. New York: Macmillan.

Wright, B. & Stone, M. (1979). *Best test design*. Chicago: MESA

Wright, B. & Stone, M. (1996). *Measurement essentials*. Wide Range. Available at www.Rasch.org

Consider joining a Rasch Listserv

ACER

To join the ACER 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 http://www2.wu-wien.ac.at/marketing/mbc/ for more information.

Rasch Workshop in San Francisco, CA

JAM Press has scheduled an Introduction to Rasch Measurement workshop for Spring 2013. This is a two-day workshop conducted by Richard Smith and Nikolas Bezruczko in San Francisco, CA. The workshop will be held on Thursday and Friday, April 25 and 26, 2013. These are the two days prior to the start of the 2013 AERA/NCME annual meeting in San Francisco. The workshop will be held at the Marines' Memorial Club and Hotel, one of the AERA hotels about two blocks from Union Square.

The Spring 2013 workshop provides an introduction to Rasch measurement, covering all aspects of dichotomous and polytomous Rasch models. The topics covered in this workshop have been revised to include several new topics that relate to large scale assessments. These include the Residual Analysis Tool, a spreadsheet application that allows the use of item/person residuals to help guide instruction and curriculum decisions in schools, decision consistency in Rasch measurement, a key concept with high-stakes assessments, the Linear Logistic Test Model developed by Gerhard Fischer, an essential topic for quality test item development, and IPARM: Item and Person Analysis with the Rasch model, an enhanced person and item analysis program that provides considerably more diagnostic information than is provides by most calibration programs.

Workshop attendees will receive a copy of three JAM Press books, Introduction to Rasch Measurement, Rasch Measurement: Advanced and Specialized Applications, and Criterion Referenced Testing: Practice Analysis to Score Reporting using Rasch Measurement, and a one year subscription to the *Journal of Applied Measurement*. The registration also includes a continental breakfast and lunch each day of the workshop. Further information on the workshop, including registration and hotel information, can be found at the JAM/JAM Press website (http://www.jampress.org) Simply click on the Rasch Measurement Workshops tab at the bottom of the page and this will take you to a pdf that contains the workshop description, agenda, and workshop registration information in a printable pdf file.

A More Accurate Pain Scale?

Pain Scales are notorious for inadequately addressing pain. Presented here is the simple, and most common, pain scale appearing in many doctor's offices.

Notable and prolific blogger, Allie Brosh, humorously points out the shortcomings of the above scale with her own interpretation of the chart:

0: Haha! I'm not wearing any pants!
2: Awesome! Someone just offered me a free hot dog!
4: Huh. I never knew that about giraffes.
6: I'm sorry about your cat, but can we talk about something else now? I'm bored.
8: The ice cream I bought barely has any cookie dough chunks in it. This is not what I expected and I am disappointed.
10: You hurt my feelings and now I'm crying!

She went on to provide a more useful chart for describing pain (see below):

0: Hi. I am not experiencing any pain at all. I don't know why I'm even here.
1: I am completely unsure whether I am experiencing pain or itching, or maybe I just have a bad taste in my mouth.
2: I probably just need a Band Aid.
3: This is distressing. I don't want this to be happening to me at all.
4: My pain is not f---ing around.
5: Why is this happening to me??
6: Ow. Okay, my pain is super legit now.
7: I see Jesus coming for me and I'm scared.
8: I am experiencing a disturbing amount of pain. I might actually be dying. Please help.
9: I am almost definitely dying.
10: I am actively being mauled by a bear.
11: Blood is going to explode out of my face at any moment.

Too Serious For Numbers: I probably have Ebola. It appears that I may also be suffering from Stigmata and/or pinkeye.

No question Brosh's pain scale is both more accurate and more useful than the simple pain scale. The traditional scale is perhaps better suited for measuring mood, as the only thing that changes is the facial expressions. This would imply that pain is entirely superficial and can be discerned primarily by evaluating one's facial expression. What's more, the categories of 0, 2, 4… 10 leave much to be desired by way of scaling, as we cannot see what a 1, 3, 5, etc. would look like. This, of course, excludes the fact that the traditional scale somewhat demeans individuals experiencing pain by asking them to accord their experience with a series of smiling and frowning faces.

Brosh's scale, on the other hand, is far more expressive, exhaustive, and better captures the experience of being in pain. Her scale includes multiple facets of pain, including body language, facial expressions, tears, blood, and a host of emotions that better describe the construct of pain. In this representation, the stick figure essentially becomes the embodiment of the pain and the textual description makes each measure much more interpretable, accurate, and relatable to those providing the evaluation. Further, Brosh addresses the many nuances of pain, including the growing intensity, extended duration, and quality. For instance, at 10, she describes pain as not only local and intense, but also as though something is being done to an individual and s/he is helpless to make it stop. In the "Too Serious for Numbers" category the pain is off the charts and is so intense that it extends beyond the physical and appears to happen at a spiritual level as well. Unlike the measures beginning at 7 that insinuate death is a real possibility and something that should be dreaded, the final category almost conveys the notion that if death does not occur one will be disappointed.

While Brosh's piece was originally intended to be humorous, it is astounding that her scale has far more elements of validity and quality measurement than many scientifically accepted scales. This affirms the notion that good measurement does not require sophisticated training in statistics or psychometrics, but rather an acceptance and understanding of the basic requirements for measurement.

Authors:

Kenneth D. Royal, University of North Carolina – Chapel Hill
Allie Brosh, Blogger at hyperboleandahalf.blogspot.com

*To access the original article see: Brosch, A. (2010). Boyfriend Doesn't Have Ebola. Probably. Available at: http://hyperboleandahalf.blogspot.com/2010/02/boyfriend-doesnt-have-ebola-probably.html.

Reliability, Separation and Strata: Percentage of Sample in Each Level

Test Reliability (Person Reliability) is routinely reported when analyzing responses to a test. It is the true variance of the sample of test respondents divided by their observed variance, where observed variance = true variance + error variance, and similarly Item Reliability can be reported for the sample of test items. Reliabilities are in the range 0 - 1, but when their values exceed 0.9, the practical implications of increases in Reliability become obscured by the range restriction. This motivated a transformation of Reliability into Separation, where Separation = square-root (true variance / error variance). Separation reports how many statistically distinguishable measurement levels exist in the sample when very high and very low measures are modeled to be accidental. A refinement of Separation is Strata, where Strata = (4 * Separation + 1) / 3. Strata models the very high and very low measures to be additional levels of performance.

For approximately normally-distributed samples, a rough estimate of the percentage of the sample in each Separation or Strata level can be computed. Levels are defined to be 3 errors apart. This distance slightly exceeds statistical significance at $p \leq .05$. The percentages in each level are shown in the Table.

John Michael Linacre

Wright BD. (1996). Reliability and separation. *Rasch Measurement Transactions, 9*(4), p. 472. Available at: www.rasch.org/rmt/rmt94n.htm

Wright BD, Masters GN. (2006). Number of Person or Item Strata: (4*Separation + 1)/3. *Rasch Measurement Transactions, 16*(3), p. 888. Available at: www.rasch.org/rmt/rmt163f.htm

Reliability	Separation levels	Percentage in each level
0.50	1	100
0.80	2	50 - 50
0.90	3	31 - 38 - 31
0.94	4	23 - 27 - 27 - 23
0.96	5	18 - 20 - 24 - 20 - 18
0.97	6	16 - 15 - 19 - 19 - 15 - 16
Reliability	Strata levels	Percentage in each level
0.20	1	100
0.61	2	50 - 50
0.80	3	23 - 54 - 23
0.88	4	14 - 36 - 36 - 14
0.92	5	10 - 13 - 34 - 13 - 10
0.95	6	8 - 16 - 26 - 26 - 16 - 8
0.96	7	7 - 11 - 20 - 24 - 20 - 11 - 7
0.97	8	6 - 9 - 15 - 20 - 20 - 15 - 9 - 6

Announcing the 2013 Joint IMEKO Symposium

The conference follows the tradition of the previous events of this well established series. There will be a special focus on measurement across physical and behavioral sciences, with the aim of high-lighting the interdisciplinary character of measurement science and of promoting constructive interactions with scientists in other disciplines.

Where: Genova, Italy
When: September 4-6, 2013

Symposium Themes include:

- Educational aspects of measurement science and instrumentation
- Real and virtual tools for education and training in measurement
- Scope and development of measurement and instrumentation as an academic discipline
- Methods and novel aspects of teachings in measurement
- Formal theories of measurement, emerging challenges and novel concepts
- Measurement in psychological, social and economic sciences
- Historical, social and ethical aspects of measurement science
- Measurement uncertainty
- Epistemology of measurement
- Mathematical modeling and design in measurement and instrumentation
- Intelligent measurement, smart sensors and virtual instrumentation
- Software engineering in measurement systems
- Artificial intelligence in measurement, computer aided measurement
- Invasive and non-invasive biological measurements
- Biological sensing modalities
- Measurement in critical care units and in clinical laboratories
- Medical imaging
- Smart sensing and smart biomedical instrumentation
- Medical information systems

For more information about the conference please visit http://www.imeko-genoa-2013.it/.

Journal of Applied Measurement
Vol. 13, No. 4, 2013

Conditional Pairwise Person Parameter Estimates in Rasch Models. *Svend Kreiner*

Examining Rating Quality in Writing Assessment: Rater Agreement, Error, and Accuracy. *Stefanie A. Wind and George Engelhard, Jr.*

Beliefs about Language Development: Construct Validity Evidence. *Mavis L. Donahue, Oiong Fu, and Everett V. Smith, Jr.*

Concurrent Validation of CHIRP, a New Instrument for Measuring Healthcare Student Attitudes towards Interdisciplinary Teamwork. *David Hollar, Cherri Hobgood, Beverly Foster, Marco Aleman, and Susan Sawning*

Using Extended Rasch Models to Assess Validity of Diagnostic Tests in the Presence of a Reference Standard. *Vivian Viallon, Emmanual Ecosse, Mounir Mesbah, Jacques Pouchot, and Joel Coste*

Measuring Work Stress among Correctional Staff: A Rasch Measurement Approach. *George E. Higgins, Richard Tewksbury, and Andrew Denney*

A Rasch Measure of Teachers' Views of Teacher-Student Relationships in the Primary School. *Natalie Leitao and Russell F. Waugh*

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

Journal of Applied Measurement
Vol. 14, No. 1, 2013

A Bootstrap Approach to Evaluating Person and Item Fit to the Rasch Model. *Edward W. Wolfe*

Using the Rasch Measurement Model to Design a Report Writing Assessment Instrument. *Wayne R. Carlson*

Using Multidimensional Rasch to Enhance Measurement Precision: Initial Results from Simulation and Empirical Studies. *Magdalena Mo Ching Mok and Kun Xu*

Using the Dichotomous Rasch Model to Analyze Polytomous Items. *Qingping He and Chris Wheadon*

With Hiccups and Bumps: The Development of a Rasch-based Instrument to Measure Elementary Students' Understanding of the Nature of Science. *Shelagh M. Peoples, Laura M. O'Dwyer, Katherine A. Shields, and Yang Wang*

Application of Single-level and Multi-level Rasch Models using the lme4 Package. *Iasonas Lamprianou*

Rasch Modeling to Assess Albanian and South African Learners' Preferences for Real-life Situations to be Used in Mathematics: A Pilot Study. *Suela Kacerja, Cyril Julie, and Said Hadjerrouit*

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

jMetrik

jMetrik is a free and open source computer program for psychometric analysis. jMetrik is available for download from www.ItemAnalysis.com. It features a user-friendly interface, integrated database, and a variety of statistical procedures. The interface is intuitive and easy to learn. It also scales to the experience of the user. New users can quickly learn to implement psychometric procedures though point-and-click menus. Experienced users can take advantage of the jMetrik command structure and write command files for executing an analysis.

Psychometric methods include classical item analysis, reliability estimation, test scaling, differential item functioning, nonparametric item response theory, Rasch measurement models, and item response theory linking and equating. New methods are added to each new version of the program.

jMetrik is a pure Java application. It runs on Windows, Max OSX, and Linux operating systems. Installation files include the needed version of Java Virtual Machine. An additional system requirement is 256MB of available memory.

AERA 2013 Rasch-related Papers
San Francisco, California
Sat., April 27 – Wed., May 1, 2013

A Comparison of Stopping Rules for Computerized Adaptive Screening Measures Using the Rating Scale Model, **Audrey Leroux, The University of Texas - Austin; *Barbara G. Dodd, The University of Texas - Austin*

Admissions to Initial Teacher Education: Are Teacher Educators' Evaluative Judgments Stable Over Time?, **Amanda K. Ferguson, University of Toronto - OISE; *Ruth A. Childs, OISE/University of Toronto; *Monique Bernadette Herbert, OISE/University of Toronto; *Olesya Falenchuk, OISE/University of Toronto*

An Empirical Study of Rating Scale Category Effects on Reliability and Validity: A Rasch Analysis, **Zongmin Kang, DePaul University*

Analyzing Teachers' Sense of Efficacy in Urban Schools (SEUS), **Mary Garner, Kennesaw State University; *Julie Kokan, Osborne High School; *Doug Hearrington, Georgia Regents University; *Marie Holbein, Kennesaw State University; *Mark Lang, Smitha Middle School; *Gita Taasoobshirazi, Kennesaw State University*

Applying Test-Equating Techniques to Identify Engaged Students, **Robert Frederick Cavanagh, Curtin University*

Assessing Science Reasoning Processes Across Topics: Challenges and Opportunities, **Ann E. Rivet, Teachers College, Columbia University; *Alison Riley Miller, Teachers College, Columbia University; *Cheryl Ann Lyons, Teachers College, Columbia University; *Mariana Schmalstig, Teachers College, Columbia University; *Kim Kastens, Columbia University*

Comparison of Models and Indices for Detecting Rater Centrality, **Edward W. Wolfe, Pearson; *Tian Song, Pearson Assessment & Information*

Development of a Work Engagement Scale: Examining the Utility of Scenario-Style Items, **Clair Marie Johnson, Boston College; *Christina Matz-Costa, Boston College; *Larry H. Ludlow, Boston College; *Melissa Brown, Boston College; *Elyssa Besen, Boston College; *Jacquelyn James, Boston College*

Development of an Instrument for Measuring Expectations for Classroom Incivility, **Luke Stanke, University of Minnesota; *Alicia Ayodele, University of Minnesota; *Doneka R. Scott, University of Minnesota*

Establishing Test Specifications for a United States Pharmacy School Equivalency Examination, **Sarah Denise Fowle, University of Illinois at Chicago*

Evaluating Parameter Recovery in the Mixture Rasch Model Based Computerized Adaptive Tests with Missing Data, **Ying-Fang Chen, University of Maryland - College Park; *Hong Jiao, University of Maryland*

Evaluation of a Behavioral Health Screening Instrument for Preschoolers Using Rasch Rating Scale Methods, **Christine DiStefano, University of South Carolina; *Fred Greer, University of South Carolina; *Jin Liu, University of South Carolina - Columbia; *Leia Kristin Cain, University of South Carolina*

Examining Erasures in a Large-Scale Assessment of Mathematics and Reading, **Aminah Perkins, Emory University; *George Engelhard, Emory University*

Examining the Language Factor in Mathematics Tests Using Explanatory Item Response Modeling, **Adnan Kan, Gazi University; *Okan Bulut, University of Minnesota - Twin Cities*

Explanatory Person-Fit Analyses with Statistical and Graphical Approaches Based on Multilevel Logistic Regression, **Angela Adrienne Walker, Emory University; *George Engelhard, Emory University*

Exploring the Relationship between Internally-Defined and Externally-Defined Academic Resilience in Mathematics, **Shanna Ricketts, Emory University; *George Engelhard, Emory University*

How Invariant and Accurate are Domain Ratings in Writing Assessment?, **Stefanie Anne Wind, Emory University; *George Engelhard, Emory University*

Increasing Measurement Precision Using a Subdimensional Item Response Model Approach, **Steffen Brandt, Kiel, Germany; *Brent M. Duckor, San José State University*

Item Parameter Drift in Computer Adaptive Testing and its Effects on Person Ability Measures, **Nicole Makas Colwell, University of Illinois at Chicago*

Job satisfaction of Canadian teachers working in privileged and disadvantaged environments, **Carla Barroso da Costa, Université de Montréal; *Nathalie Loye, University of Montreal*

Measuring Coping Resources for Stress Management: A Rasch Analysis, **Jennifer Mellott, Kent State University; *Philip Gnilka, DePaul University*

Measuring Perfectionism With the Almost Perfect Scale—Revised (APS-R): A Rasch Analysis, **Edward C Bolden, Kent State University; *Jennifer Mellott, Kent State University; *Philip Gnilka, DePaul University*

Measuring Student Growth Through Rasch Vertical Equating: Learning Within and Across Grades, **Gregory E. Stone, University of Toledo; *Kristin L.K. Koskey, The University of Akron; *Toni A. Sondergeld, Bowling Green State University; *MingYang Liu, University of Toledo*

Program Leadership in Western Australian Secondary Schools: A Rasch Model Investigation of Incumbents' Perceptions, **Robert Frederick Cavanagh, Curtin University; *Zoe A Brooks, Curtin University*

Self-Efficacy in Math: A Rasch Measurement Approach, **Kris Aric Knisely, Emory University; *Michael Nguyen, Emory University; *Mei-Lin Chang, Emory University*

The Effects of Mixture Distribution of Calibration Sample on Accuracy of Rasch Item Parameter Estimation in Computerized Adaptive Test, **Shudong Wang, NWEA*

The Rasch Model Plus Ability-Based Slipping, **Chao Xie, University of Maryland - College Park; *Hong Jiao, University of Maryland*

Using Rasch Modeling to Support Validation of a Developmental Progression for Area, **Jeffrey E. Barrett, Illinois State University; *Douglas W. Van Dine, University of Denver; *Craig Cullen, Illinois State University; *Cheryl L. Eames, Illinois State University; *Melike Kara, Illinois State University; *Amanda Miller, Illinois State University*

Validating the Internal Structure of the Performance Assessment for California Teachers (PACT): A Multidimensional Item Response Model Study, **Brent M. Duckor, San José State University; *Katherine Furgol Castellano, University of California - Berkeley; *Kip T. Tellez, University of California - Santa Cruz; *Mark R. Wilson, University of California - Berkeley*

Validation of a Developmental Progression for Volume Using Rasch Modeling, **Julie Sarama, University of Denver; *Douglas W. Van Dine, University of Denver*

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 June 1, 2013, so please make your submission by May 1, 2013 for full consideration. Please email Editor\at/Rasch.org with your submissions and/or ideas for future content.

Rasch-related Coming Events

Mar. 8, 2013, Fri. UK Rasch User Group Meeting, Manchester, UK, www.rasch.org.uk

Mar. 25-27, 2013, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Apr. 4, 2013, Thurs. MESA (UIC) Online programs webinar. 12 p.m. and 6 p.m. CST (E. Smith), http://www.uic.edu/scs/education/degree/online-mesa/webinar.html

Apr. 25-26, 2013, Thurs.-Fri. In-person workshop: Introduction to Rasch Measurement (R. Smith, N. Bezruczko), San Francisco, CA, www.jampress.org

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,

May 31-June 28, 2013, Fri.-Fri. Online workshop: Practical Rasch Measurement – Core Topics (E. Smith, Winsteps), www.statistics.com

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

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.