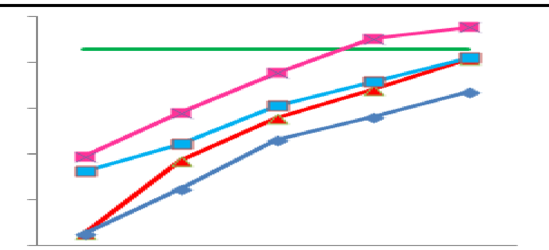
Lexile Reading Levels



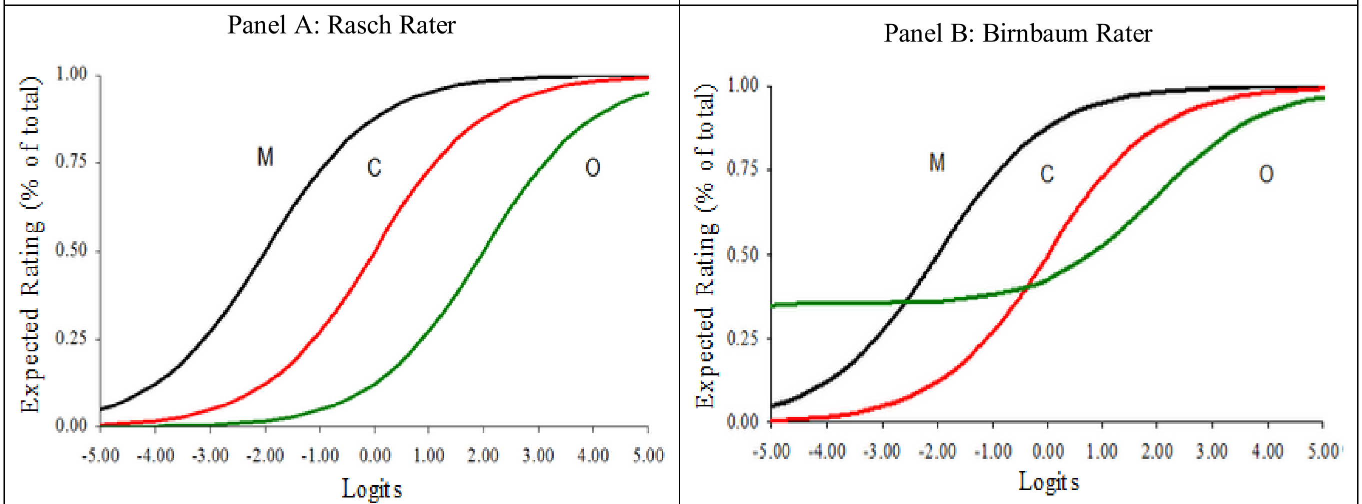
RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 25 No. 2 | Autumn 2011 | ISSN 1051-0796

Rater-Mediated Domain Response Functions

Rater Invariant Measurement	Rater Variant Measurement



Panel C

Domain			
Hard	O	O	O
Medium	C	C	C
Easy	M	M	M
Writing Proficiency:	Low	Medium	High

Panel D

Domain			
Hard	C	C	O
Medium	M	O	C
Easy	O	M	M
Writing Proficiency:	Low	Medium	High

Figure 1 Impact of Crossing Rater-Mediated Domain Response Functions.
The domains are Mechanics (M), Content (C), and Organization (O).

Rater-mediated assessments are used extensively in a variety of educational contexts (Engelhard, 2002). In evaluating the quality of ratings obtained in these contexts, the idea of rater-mediated operating characteristic functions (rm-OCFs) has not been systematically explored. OCFs can be used to enhance the substantive interpretations of rater behaviors. For example, the substantive interpretation of crossing item response functions (IRFs) is fairly well known (Wright, 1997), and Perkins and Engelhard (2009) have discussed crossing person response functions (PRFs). Similar ideas can be used to develop rater-mediated domain response functions (rm-DRFs), as well as rater-mediated person response functions (rm-PRFs). Just as crossing IRFs or PRFs create differential ordering of item difficulty and person performance, crossing rm-DRFs and rm-PRFs have implications for the substantive interpretation of rater behavior. When rm-DRFs cross, the interpretation of the domains across the latent variable is not invariant

Table of Contents

above and below the intersection points. This note provides an illustration of crossing rm-DRFs, and demonstrates the substantive interpretation of this situation.

Both Rasch (1960/1980) and Birnbaum (1968) propose operating characteristic functions for dichotomous responses that can be used to model dichotomous ratings. For example, a Rasch model for dichotomous ratings can be written as follows:

$$\phi_{nmi} = \frac{\exp(\theta_n - \lambda_m - \delta_i)}{1 + \exp(\theta_n - \lambda_m - \delta_i)} \quad [1]$$

where φ_{nmi} is the probability, θ_n is the judged location of person n on the latent variable (e.g., writing proficiency) by rater m with a severity of λ_m on domain i with a judged difficulty of δ_i.

A Birnbaum Model for dichotomous ratings can be written as

$$\phi_{nmi} = c_i + (1 - c_i)\frac{\exp(\alpha_i(\theta_n - \lambda_m - \delta_i))}{1 + \exp(\alpha_i(\theta_n - \lambda_m - \delta_i))} \quad [2]$$

where α_i is a scale parameter that varies across domains, and c_i is the lower asymptote of the function that represents rater reluctance to assign low ratings to persons (a comparable upper asymptote can also be introduced for rater reluctance to assign high scores).

In the context of rater-mediated assessments, the rm-DRF for a Rasch rater (λ_R) on domain one (δ_1) rated dichotomously (fail/pass) can be written as:

$$\phi_{nR1} = \frac{\exp(\theta_n - \lambda_R - \delta_1)}{1 + \exp(\theta_n - \lambda_R - \delta_1)} \quad [3]$$

and for a Birnbaum rater (λ_B):

$$\phi_{nB1} = c_1 + (1 - c_1)\frac{\exp(\alpha_1(\theta_n - \lambda_B - \delta_1))}{1 + \exp(\alpha_1(\theta_n - \lambda_B - \delta_1))} \quad [4]$$

The general requirements for invariant measurement are summarized by Engelhard and Perkins (2011), and these requirements can be extended for raters (Wind & Engelhard, 2011):

- The measurement of persons must be independent of the particular raters that happen to be used for measuring: *Rater-invariant measurement of persons.*

Figure 1 illustrates the effects of crossing rm-DRFs for two raters who are rating writing proficiency using three domains: Mechanics (M), Content (C), and Organization (O). Panel A is a Rasch rater with non-crossing DRFs, while Panel B is a Birnbaum rater with crossing DRFs. Panel C shows a substantive interpretation for non-crossing DRFs that produce comparable judged domain difficulties over subgroups of persons. The ordering of the three domains is invariant with the mechanics (M) domain judged easiest and organization (O) domain judged as hardest across the latent variable of writing proficiency. Non-crossing DRFs result in equivalent ordering of domains across subsets of persons, and yields invariant measurement from the Rasch rater.

Panel D shows the substantive interpretation of crossing DRFs based on a Birnbaum rater. The meaning of person performance on domains varies as a function of person subgroup locations on the latent variable of writing proficiency. The Rasch rater interprets the domains in a comparable way over subgroups with domains ordered as M < C < O, while the domain difficulties are variant for the Birnbaum rater. The Birnbaum rater rates the organization (O) domain easiest for persons with low writing proficiency, while organization (O) is rated hardest for persons with high writing proficient.

In practice, model-data fit and the requirements of invariant measurement can be usefully visualized with OCFs. This note highlights the need for researchers to examine differential domain functioning as an additional aspect of model-data fit within the context of rater-mediated assessments. It is recognized that domains may function differently over subgroups of persons (differential domain functioning).

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

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability, Part 5. In F.M. Lord and M.R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Engelhard, G. (2002). Monitoring raters in performance assessments. In G. Tindal and T. Haladyna (Eds.), *Large-scale Assessment Programs for ALL Students: Development, Implementation, and Analysis,* (pp. 261-287). Mahwah, NJ: Erlbaum.

Engelhard, G, & Perkins, A.F. (2011). Person response functions and the definition of units in the social sciences. *Measurement: Interdisciplinary Research and Perspectives*, *9*, 40-45.

Perkins, A., & Engelhard, G. (2009). Crossing person response functions. *Rasch Measurement Transaction, 23*(1), 1183-1184.

Rasch (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. (Expanded edition, Chicago: University of Chicago Press, 1980).

Wind, S.A., & Engelhard, G. (2011, July). *Evaluating the quality of ratings in writing assessment: Rater agreement, precision, and accuracy*. Paper presented at the Pacific Rim Objective Measurement Seminar (PROMS) in Singapore.

Wright, B.D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice*, Winter, 33- 45, 52.

Rasch Measurement SIG

Nominations for Officers

Dear Rasch SIG members,

I am writing to provide you with information concerning the 2011 AERA General Election. Kenneth Royal and I will complete our terms as Rasch SIG Secretary/Treasurer and Chair, respectively, at the 2012 Annual Meeting. Because all SIG elections are now incorporated into the AERA General Election, and that process requires me to complete and submit a form for each position to be contested in the 2011 election by November 15, 2011, I am beginning the nominations process now.

Relevant information about the process is shown toward the end of this message. I encourage you to nominate (or self-nominate) someone who you think would be a good Chair or Secretary/Treasurer of the SIG.

Please email, to *michael.j.young \at/ pearson.com* your (self-)nominations for the offices of Chair and Secretary/Treasurer prior to November 1st, 2011.

Please include the individual's name, contact information, and the position for which that person is being nominated. I will contact those who are nominated to confirm that they are willing to serve and to request a candidate statement prior to the November 15th deadline for submitting nominations to AERA.

The relevant sections of the SIG By-Laws, in full at www.raschsig.org/bylaws.doc contain the following points:

- There are two elected positions: Chair and Secretary/Treasurer.
- Elections take place via email balloting of the Rasch Measurement SIG members 3 months prior to the annual meeting.
- All SIG members are eligible to serve as officers.
- The term of each office is 2 years, commencing and expiring at the Annual AERA Meeting.
- No person shall serve more than 2 consecutive terms in a single office.
- This call for nominations is to be distributed electronically and published in the newsletter.
- The Chair shall be responsible for the general administration of the Rasch SIG and act as liaison between the SIG and AERA, shall preside at all meetings of the Executive Committee and at the annual business meeting, and shall appoint ad hoc committees as needed.
- The Secretary/Treasurer shall be responsible for the safe keeping of all financial documents and any official correspondence and meeting minutes of the Rasch SIG, will be responsible for maintaining the Rasch SIG website (www.raschsig.org) or appointing an appropriate representative as needed.

Fifth International Conference on Probabilistic Models for Measurement

January 23-25, 2012
Perth, Australia

The University of Western Australia hosts the Fifth International Conference on Probabilistic Models for Measurement in Education, Psychology, Social Science and Health from Monday, 23 January 2012 to Wednesday, 25 January 2012.

The conference is preceded by two weeks of courses on social measurement, in particular Rasch measurement theory and practice, featuring the RUMM2030 software package (January 9-15 and January 16-20. Details at :

www.education.uwa.edu.au/raschconference

This Conference is followed by a Pearson Global Research Conference, "The Role of Technology and Assessment in System-wide Improvement" (January 27-28). Details at:
www.pearson.com.au/marketing/corporate/pearson_global

Rasch Measurement Transactions (RMT)

Nominations for Editor

RMT is a quarterly publication of the Rasch Measurement SIG. It contains announcements of coming SIG and related activities, historical records of those activities, research notes and other information likely to be of interest to Rasch practitioners. It is free to SIG members and to the general public. The Editor of RMT is appointed by the elected SIG officers. A new Editor will be appointed during the SIG Business Session at the 2012 AERA Annual Meeting. On the current schedule, the new Editor's first RMT will be published in June 2012.

The SIG Officers are looking for a Rasch enthusiast who can creatively develop RMT in the current fast-changing Internet environment. Expected skills include proficiency in the theory and practice of Rasch, competence in word-processing and web-publishing, and especially the ability to discover relevant authors and source-material.

Please email, to *michael.j.young \at/ pearson.com* your (self-)nominations for Editorship of RMT prior to December 1st, 2011.

Previous RMT editors are Richard M. Smith (1987-1989) and John "Mike" Linacre (1989-2012).

Editorship of RMT is an exciting opportunity for a dynamic researcher eager to influence the development of Rasch methodology and the advance of science in general.

Is Combining Samples Productive? A Quick Check via Tests of DIF

Questions have recently been asked about combining samples from different populations to obtain more precise estimates of Rasch (1960) model parameters. *Ceteris paribus,* the more data that is available for a given test, the more precise the parameters will be. There can be times, however, when combining different sets of test data may be problematic. Fortunately, a simple way of checking for problems with aggregating data exists. The different samples can be entered into a Rasch analysis as person factors/ facets and the items checked for Differential Item Functioning (DIF).

Application of the Rasch model assumes that parameters are invariant with respect to populations. The presence of DIF voids this assumption. A test in which many items suffer from DIF will produce person ability estimates that are biased. If the DIF is "non-uniform" (e.g., "Sample A's" and "Sample B's" item response functions intersect), then there is a problem and the data should not be combined. In cases of uniform DIF, the item response functions do not intersect, which means that a mathematical transformation could render these curves parallel.

Uniform DIF can be treated in a very powerful way in RUMM2030 by "splitting" the item. This means the Rasch model is used to calculate two different item difficulty parameters for an item affected by uniform DIF – one for Sample A examinees and one for those from Sample B. When RUMM2030 calculates person ability estimates, then depending upon the examinee's classification, one of the two item difficulty estimates is used. For example, for a Sample A examinee, the calculation of that person's ability will use the Rasch item difficulty parameter estimate for Sample A examinees for that item. A simple t test upon the two sets of person ability scale scores (i.e., split and unsplit) can reveal if the mean person ability estimates are statistically significantly different.

I have encountered two recent problems concerning the combination of sample sizes and DIF. The first concerned a vocabulary test consisting of 104 dichotomous items. Initial Rasch calibration was conducted using the RUMM2030 program on a sample of 510 readers. Of these, 288 were classified as an "English Learner" and 222 were "English Proficient". One hundred and seventy three participants were in Grade 4 and 334 were in Grade 3 at the time of test administration.

Overall fit of the Rasch model to the data was poor as many items did not fit the Rasch model. The Total Item Chi Square, which is a sum of individual item chi squares, was 1,137; df = 312, $p < .001$. The PSI reliability, however, was quite high, being at .96.

The test developers thought that the test suffered from "multidimensionality", but performing a principle components analysis on residual correlations did not reveal any instance of this. DIF was investigated in RUMM2030 by the calculation of Item Response Curves (ICCs) for each person factor for each item. If DIF is not present in the data for an item, there will be no discernable differences between person factor ICCs for that item. Additionally, main effects for the person factor in ANOVA analyses of item residuals will not be statistically significant.

There was no DIF when the person factor involved was Grade. Perhaps not surprisingly, there was a serious and substantial amount of DIF when the English Proficiency/Learning factor was assessed. A DIF analysis of Item 77 is displayed in Figure 1. Two ICCs have been calculated – one for participants who were classed as English Proficient (blue ICC) and one for English Learners (red ICC). If there was no difference between these two groups in performance on this item, then both the red and blue ICCs would both fall on the theoretical grey ICC. In this case they do not, and so therefore this item suffers from DIF.

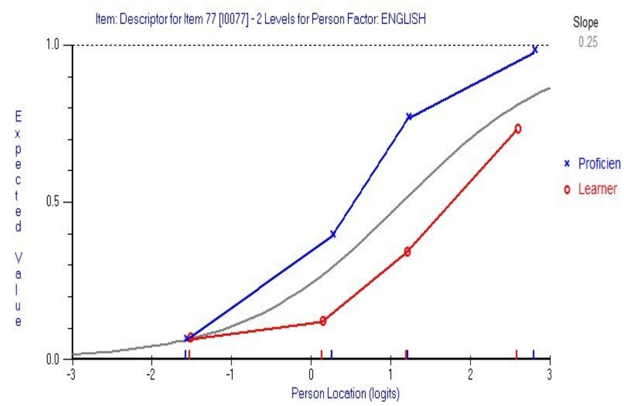


Figure 1. DIF analysis of item 77 of the vocabulary test.

Item 77 was amongst those items split. Figure 2 represents the "split" Item 77 for English Learners. The difficulty of Item 77 for English Learners was 2.095. In the original unsplit item, the difficulty was 1.143. Hence the split Item 77 for English Learners, which fits the Rasch model, is a more difficult item for these examinees than the original, which did not fit the Rasch model.

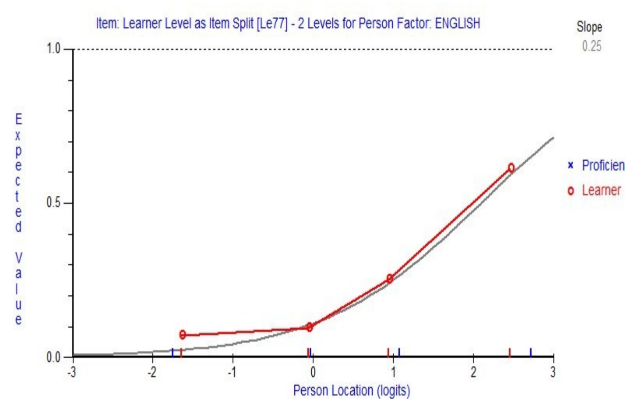


Figure 2. Item "Le77" created by splitting item 77.

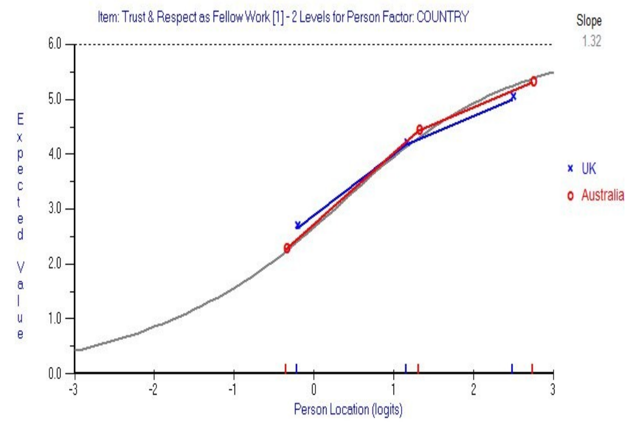


Figure 3. A DIF analysis of the first item of the interpersonal trust questionnaire.

Almost all instances of English Proficiency/Learning DIF in the test were uniform, meaning that most DIF items were split (34 items in total). The test was then recalibrated and all misfitting items, both split and unsplit, were removed from analysis. Forty six items in total were removed. This substantially improved the overall fit of the Rasch model to the data (chi square = 356, df = 276, $p < .001$). Whilst still statistically significant, the magnitude of the overall chi square statistic was reduced by more than two thirds. The PSI reliability coefficient was .94, which meant that test reliability was only marginally affected by the removal of items. To test the difference between calibrations, person ability estimates from the initial and final calibrations were obtained and a paired samples t – test was conducted. The difference between the means of .315 logit was statistically significant (t(506) = 23.82, $p < .001$, one tailed). Hence the DIF in the initial calibration caused person ability estimates to be biased by an average of almost one third of a logit.

Thus "multidimensionality" was not the culprit for poor fit of the Rasch model. It was the test developers' decision to administer the test to samples of examinees from two very different populations – those just beginning to learn English and those who were proficient in it. Nonetheless, item splitting salvaged the test calibration.

The other problem was something quite different. An academic colleague combined two samples of managers – 107 from the U.K. and 85 from Australia – to analyze a questionnaire of interpersonal trust. In response to a paper written on the project, a reviewer stated that "…combining the UK and Australian samples of sales managers into one dataset generates additional confounding… country-level effects will potentially bias the estimates and this poses a serious problem". Testing items for DIF was a means of being able to test the reviewer's conjecture.

Figure 3 displays the UK and Australian sample ICCs for the first test item, which read "*Most people, even those who aren't close friends of the marketing manager, trust and respect him/her as a fellow worker.*"

Like the first item in Figure 3, no other item in the test suffered from DIF. Moreover, the test was reliable (Cronbach's alpha = .94). The combining of samples from different nationalities was therefore justified as this caused no discernable bias in the Rasch parameter estimates.

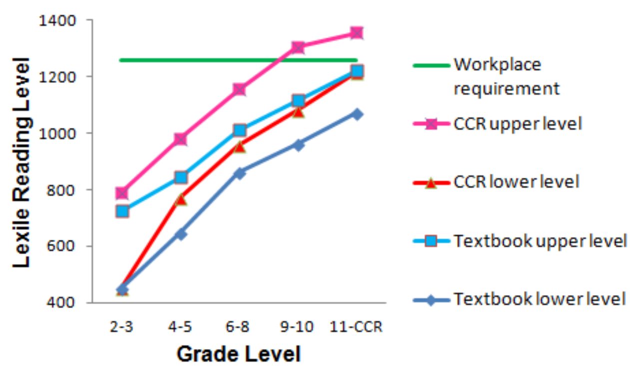
Andrew Kyngdon, MetaMetrics, Inc.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

RUMM Laboratory Pty Ltd. (2009). *RUMM2030*™ [Computer software]. Perth, Australia: Author.

Lexiles Demystify the Reading Standards

The *Common Core State Standards for Reading* (2010) propose that the difficulty of reading material at all grade-levels in US schools should be raised in order that graduating students should attain *"College and Career Readiness"* (CCR) reading skills. A simple plot (based on Smith, 2011) shows the Lexile standard required for Workplace reading proficiency (green line), the reading levels of current classroom texts (between the bluer lines) and the proposed levels (between the redder lines). Notice that the biggest advance is in the lower grades. Children must be able to read early and well.



National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects. Washington, DC: Author.

Smith, Malbert III (2011) Bridging the Readiness Gap: Demystifying Required Reading Levels for Postsecondary Pursuits. MetaMetrics® Policy Brief
lexile.com/m/uploads/downloadablepdfs/Policy_Brief_1_Bridging_the_Readiness_Gap.pdf

Psychometric Models and Methods in R

Much psychometric functionality is already contained in base R and there is considerable overlap between tools for psychometry and tools described in other places, particularly in SocialSciences. We give a brief overview of R packages that are closely related to psychometric methodology.

cran.r-project.org/web/views/Psychometrics.html

Patrick Mair and Reinhold Hatzinger

Considering Large Group Differences in Ability in DIF Analysis

Differential item functioning (DIF) occurs when an item has a different probability of endorsement for different groups of respondents who are equivalent on the measure. When groups are not equivalent, matching of respondents on ability or other method of controlling for group differences is necessary. Nevertheless, there is evidence that some DIF procedures may result in increased false positive DIF results when groups differ in ability (DeMars, 2010; Li & Stout, 1996). Previous studies have generally focused on the impact of small to moderate (e.g., 0.5 to 1.0 logits) group differences. What effect do larger group differences have on DIF results?

A series of Monte Carlo simulations were performed to answer this question. Dichotomous item responses to a 25-item instrument were simulated according to the Rasch model for two groups of simulees (n = 250 or 500 per group). In each dataset, a common set of item calibrations derived from a uniform random distribution were used for both groups (i.e., no DIF was simulated). Group means for both reference and focal groups ranged from -1.5 to 1.5 logits in increments of 0.5 logits (group standard deviations = 1.0). For each combination of sample size, focal group mean, and reference group mean, 100 datasets were generated. False-positive DIF using both t-test comparisons of item calibrations and the Mantel-Haenszel (M-H) test was based on statistical significance at the .05 level. Simulations were performed using R (version 2.13), the rWinsteps package (version 1.01) and Winsteps (version 3.72.3).

The figure shows the false positive DIF rates for each method by focal group – reference group differences in ability and sample size. Both methods resulted in false positive DIF rates that are generally within the nominal .05 level, with M-H having lower error rates, particularly with absolute group differences ≥ 2 logits. Only with the t-test procedure, when the number of respondents per group equaled 500 and when absolute group differences were ≥ 2.5 logits did the false positive rate exceed 5 percent.

Conclusions: These results suggest that **DIF methods commonly used in conjunction with Rasch measurement are robust against large differences in group ability**. The Mantel-Haenszel procedure resulted in a lower false positive rate and may be the more appropriate method when group differences and sample sizes are large. A limitation of the analysis is that power in detecting true DIF was not assessed.

Barth Riley

DeMars, C. E. (2010). Type I error inflation for detecting DIF in the presence of impact. *Educational and Psychological Measurement, 70(6),* 961-972.

Li, H.-H., & Stout, W. F. (1996). A new procedure for detection of crossing DIF. *Psychometrika, 61(4),* 647-677.

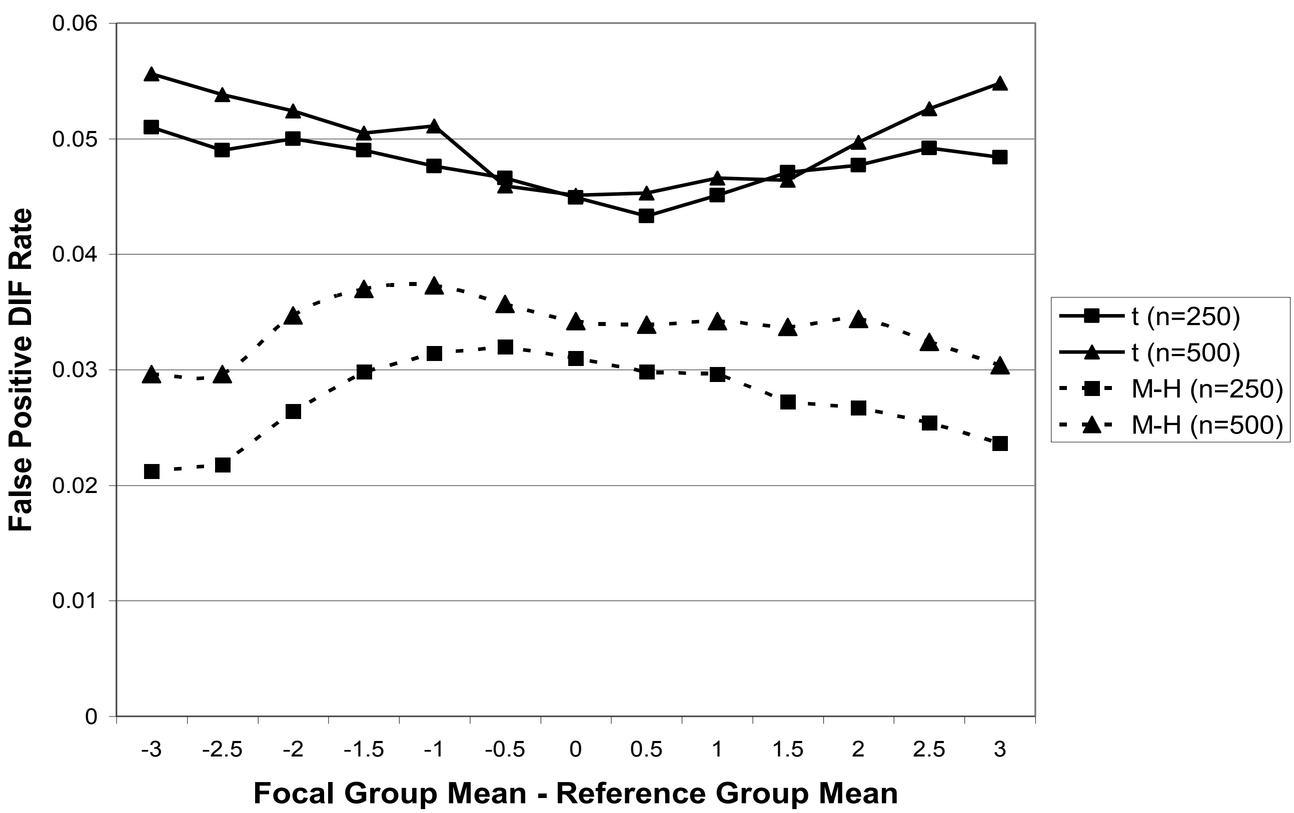


Figure. False Positive DIF Rate by Group Mean Difference and Sample Size.

The Future of Computer Science (Rasch Measurement?)

Peter Lee, the Managing Director of Microsoft Research Redmond, has made some comments in an interview (Knies, 2011) about the future of Computer Science. These comments also well express the future of Rasch Measurement.

Lee: "The number of potential and real breakthroughs [in computing] waiting to have tremendous impact on people's lives is huge. It has just exploded."

> For Rasch, we are still at the waiting stage. We have not yet reached the explosion.

"... advances in machine learning and the manipulation of massive datasets figure to transform the things computers can do and what can be done with them."

> Rasch measurement is also being transformed as decision-makers try to make sense of ever larger and more complex databases. The need to impose the conceptual order of unidimensional variables upon chaotic data becomes increasingly pressing.

"...the first decade for such organizations [the great industrial research labs] is all about the ability to recruit great personnel, which leads to a reputation as a legitimate force in the research community."

> Though not concentrated in a single organization, the Rasch community at large has, in fact, had 50 years of recruiting great personnel and has cemented its reputation as a legitimate force in the world of research and practice. Might we say that the recruitment and reputation stage peaked around 1990? By that time, the basic array of models, estimation methods, fit assessments, software, and practical examples were in place and moving into routine use.

"The second decade is concerned with growth, among individual researchers and as a lab."

> Again, though not involving a single organization, growth in the number of researchers building on the work of the founding innovators has been significant. My recent search on Google Scholar for "Rasch model OR analysis OR scale OR scaling OR measurement" in *article titles* gives 2,510 hits. Breaking that down by years, 1961-1970 has 10 hits, 1971-1980, 200; 1981-1990, 283; 1991-2000, 496, 2001-2010, 1,160, and 109 articles have come out so far this year. These certainly underestimate the actual numbers of Rasch articles published, as the majority do not include "Rasch" in their titles. Though the numbers here may be only a sample of the actual totals, the growth trend is evident. See also "The Appearance of Rasch in Journal Articles", RMT 24:4, 1311.

"Once a certain critical mass is achieved, then, in the third decade, if the investment has been well-managed and sustained, people begin to turn their sights to a legacy that extends beyond individual fame and respect to a more meaningful impact with the potential to effect a significant societal difference."

> Rasch has not yet reached the "third decade". Taking the period to 1990 as the "first decade," does the "second decade" span 1990 to 2000, or 1990 to 2010, or something in between? Though there is great potential for Rasch measurement to effect a significant societal difference, there is very little evidence of many turning their sights in that direction. Various individual articles have touted item banking, adaptive instrument administration, or construct theory as transformative technologies, but none of these explicitly explore a wide range of social impacts.
>
> The time is ripe for such "third decade" explorations. Paraphrasing Peter Lee by replacing "computer-science" with "Rasch measurement" gives us:

"What we in *Rasch measurement need* is the motivation to break out of a mindset of thinking inwardly about understanding *measurement* and the mechanisms of *measurement* and instead look outward to the role of *measurement* in the world. The success of *measurement* research is expanding rapidly, but for all of that dramatic progress and expansion, the gap that people see between the progress of *measurement* science and what society needs from *measurement* .... People still see that gap not shrinking."

> I couldn't have said it better myself. However, all of us in Rasch measurement can go further than what Lee said in response to the last question in the interview. The question was "How, then, can the value of research be measured?"

"We don't do it by dollars. What I ask from managers is to think about impact: 'What high-impact results have you produced? Give me something to brag about. Tell me how you're affecting Microsoft's ... businesses. Tell me how you're contributing to the research community. What is your plan? How are you structuring your team to position itself to make progress?"

> Lee rightly eschews financial metrics, instead talking in terms of qualitative indicators of impact, but that's as far as he gets. His response immediately suggests to us that Microsoft needs to construct a latent variable of "research value". Lee has identified some of the item content as well as the "bragging" rating-scale on which items could be scored. Perhaps someone reading this issue of RMT will calibrate a practical quantitative measurement tool for "research value" able to support the decision-making needs of Microsoft and other research-funding agencies.

William P. Fisher, Jr.

Knies R. (2011) 20 Years On, a Future Brighter than Ever. Microsoft Research.
research.microsoft.com/en-us/news/features/anniversaryoverview-092611.aspx

Francis Galton – Ahead of His Time

"The senior wranglers [highest scoring students at Cambridge University] had more than thirty, or thirty-two times the ability of the lowest men on the lists of honours. They would be able to grapple with problems more than thirty-two times as difficult; or when dealing with subjects of the same difficulty, but intelligible to all, would comprehend them more rapidly in perhaps the square root of that proportion." (Francis Galton, "Hereditary Genius", 1892, p. 61).

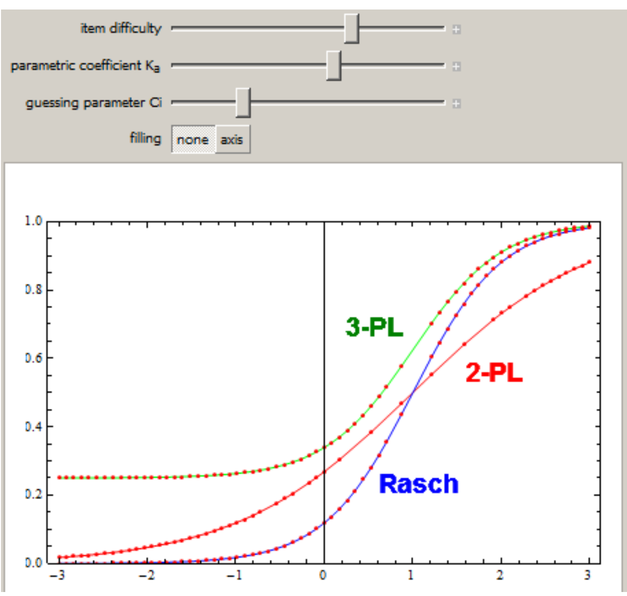
Galton has grasped the concept of separability. The ability of the students is conceptualized separately from the difficulty of any particular set of items, and *vice versa.*

Understanding Galton's statements in terms of the multiplicative Rasch model, "32 times" = 3.5 logits. In many educational contexts, 1 logit corresponds to one year's growth. So it makes sense that the best-performing students were about 4 years ahead of the worst-performing students.

Andrew Stephanou and colleague

Comparing Rasch and IRT ICCs (IRFs)

Wolfram Demonstrations Project provides a playful pictorial comparison of the shapes of Rasch, 2-PL and 3-PL item characteristic curves (item response functions). demonstrations.wolfram.com/123ParameterLogisticRaschAndBirnbaumModelsAndItemAnalysis



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

SIG Chair: Michael Young
Secretary: Kenneth Royal
Program Chairs: Leigh Harrell & Stephen Jirka
SIG website: www.raschsig.org

Rasch-related Coming Events

Oct. 21 - Nov. 19, 2011, Fri.-Fri. Online course: Rasch - Core Topics (Linacre) www.statistics.com

Nov. 25, 2011, Fri.. VI Workshop sobre Modelos de Rasch en Administración de Empresas. Nuevas Tendencias. Tenerife, Spain. www.institutos.ull.es

Nov. 30 - Dec 2, 2011, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), UK

Dec. 5-7, 2011, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Jan. 9 - Apr. 27, 2012, Mon.-Fri. Online course: Rating Scale and Questionnaire Design and Analysis (E.V. Smith), education.uic.edu

Jan. 9-15, 2012, Mon.-Wed. In-person workshop: Introductory Rasch course (Andrich, RUMM2030),

Jan. 16-20, 2012, Mon.-Wed. In-person workshop: Advanced Rasch course (Andrich, RUMM2030), Perth, Australia, www.education.uwa.edu.au

Jan. 23-25, 2012, Mon.-Wed. Fifth International Conference on Probabilistic Models for Measurement, Perth, Australia, www.education.uwa.edu.au

March 20, 2012, Tues. 6th UK Rasch User Group Meeting, Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric/News1.htm

March 21-23, 2012, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Apr. 11-12, 2012, Wed.-Thurs. IOMW International Objective Measurement Workshop, Vancouver BC, Canada, Announcement

Apr. 13-17, 2012, Fri.-Tues. AERA Annual Meeting, Vancouver BC, Canada, www.aera.net

May 23-25, 2012, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

May 28-30, 2012, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Aug. 6-9, 2012, Mon.-Thur. PROMS2012, Jiaxing University, Zhejiang Province, P.R.China, Facebook

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

Sept. 13-14, 2012, Thurs.-Fri. In-person workshop: Advanced Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric