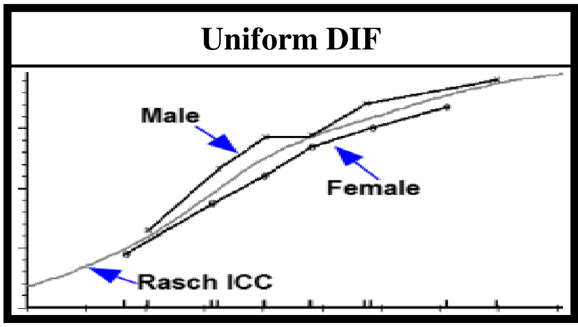
Uniform DIF



RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 20 No. 4 Spring 2007 ISSN 1051-0796

Applying The Rasch Rating-Scale Model To Set Multiple Cut-Offs

Setting multiple cut-offs in educational contexts where score points are required to indicate transitions from one ability level to the next is a challenging issue. One such context is the Common European Framework of Reference for Languages (CEF). The CEF is a six-point proficiency scale with descriptors for each band in the form of 'can-do statements'. The levels on the CEF are A1, A2, B1, B2, C1 and C2, A1 being the lowest level and C2 the highest. Linking tests to the CEF and validating the claims of links to the CEF is an important issue that European language testers are wrestling with. In this paper a methodology for linking tests to the CEF or any other similar proficiency scale is suggested.

Procedures

The material was a reading comprehension test comprising 75 items. A group of 20 raters who were familiar with the CEF proficiency scale and its descriptors were asked to rate the 75 items, indicating what minimum ability level on the six-point CEF proficiency scale a student should exhibit to get each item right. The items were rated from 1 to 6, 1 indicating the lowest CEF level and 6 the highest. The items were then calibrated on the basis of these ratings using Andrich's (1978) rating scale model. The next step was to calibrate the items on the basis of actual student performances. The following tables show the descriptive statistics for the item measures based on the two analyses.

Table 1. Item measure summary.

Item measure summary statistics	Rater-based analysis	Student-based analysis
N	75	75
Mean	.26	-.00
Median	.16	-.05
Std. Deviation	2.72	1.80
Range	10.40	8.42
Minimum	-5.97	-3.28
Maximum	4.43	5.14
Reference difficulty	0.00	item mean

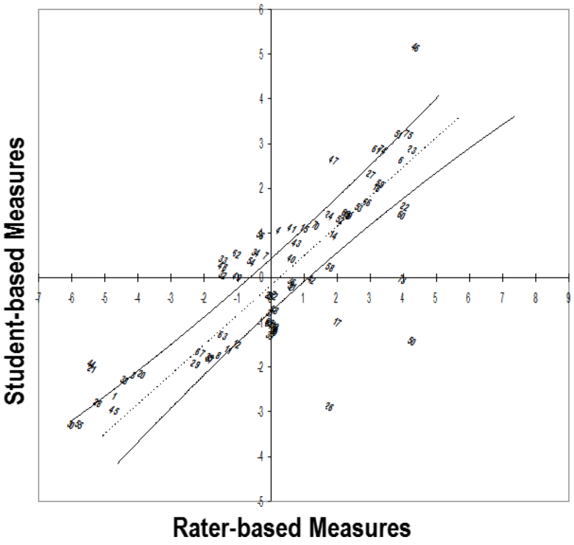


Figure 1. Cross-plot of item measures from rater-based and student-based analyses

The cross plot of the item calibrations from the two analyses is shown in Figure 1. The two sets of item calibrations, i.e., those based on raters and those based on the students' performances, correlated at 0.80. It can be seen that there are a few conspicuous outliers, and there may be two trendlines, one for the upper half of the plot, and the other for the lower, but the overall pattern is clear. The slope of an empirical joint "best fit" line (through the two means, and two means+1 S.D.) is 0.66. The mean difference between the average item measures is 0.26 logits. Thus the person measures were converted into the rater frame-of-reference by means of the equating formula:

$$M2 = (M1 - mean(1))*SD(2)/SD(1) + mean(2)$$
$$\text{i.e., Adjusted measure} = (measure-.00)/0.66 + 0.26$$

Table of Contents

Rater analysis equated with student analysis

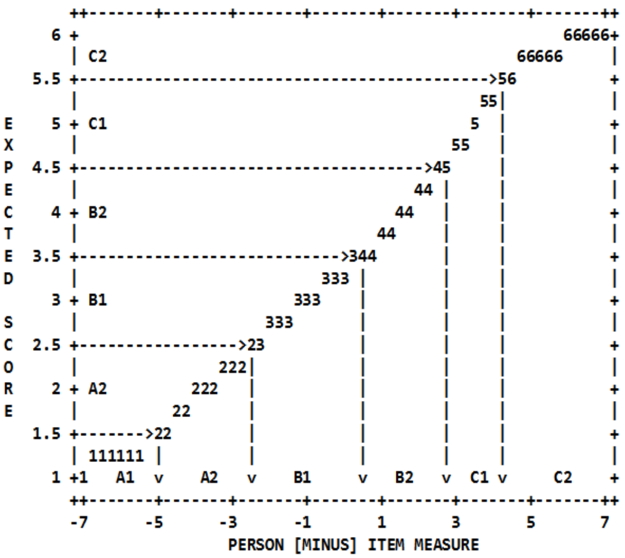
When the person measures are equated for both the intercept and the slope of the trendline, they are mapped into the framework of the rater-based analysis. Table 2 shows the descriptive statistics for the 160 person measures in three different modes: (1) unequated, (2) equated with the rater-based analysis by correction for intercept only, and (3) equated with the rater-based analysis by correction for both intercept and slope.

Table 2: Descriptive statistics for 160 persons in three different modes.

	Person Measures Unequated	Person Measures Equated for Intercept	Person Measures Equated for Intercept & Slope
N	160	160	160
Mean	.07	.33	.37
Median	.26	.52	.65
Mode	-1.51	-1.25	-2.02
Std. Deviation	1.34	1.34	2.03
Range	5.78	5.78	8.73
Minimum	-3.28	-3.02	-4.70
Maximum	2.50	2.76	4.04

Setting cut-points

Half-score-point thresholds on the reference item at zero logits in the rater analysis set the cut-off scores. Since the person measures have been brought to the framework of the rater-based analysis these half-score-point thresholds are directly applicable to the person measures after equating. The expected score ICC for the reference item is shown in Figure 2. The half-score point intervals are indicated on the latent variable. The locations of the 6 proficiency levels are indicated by their codes, A1, etc.



```
       ++-------+-------+-------+-------+-------+-------+-------++
     6 +                                                    66666+
       | C2                                             66666     |
   5.5 +--------------------------------------------->56          +
       |                                           55|           |
E    5 + C1                                       5  |           +
X      |                                        55   |           |
P  4.5 +------------------------------------->45     |           +
E      |                                    44 |     |           |
C    4 + B2                               44   |     |           +
T      |                                44     |     |           |
E  3.5 +---------------------------->344       |     |           +
D      |                          333 |        |     |           |
     3 + B1                    333    |        |     |           +
S      |                    333       |        |     |           |
C  2.5 +----------------->23          |        |     |           +
O      |               222|           |        |     |           |
R    2 + A2         222   |           |        |     |           +
E      |          22      |           |        |     |           |
   1.5 +------->22        |           |        |     |           +
       | 111111 |         |           |        |     |           |
     1 +1   A1  v    A2   v    B1     v   B2   v  C1 v     C2    +
       ++-------+-------+-------+-------+-------+-------+-------++
       -7      -5      -3      -1       1       3       5       7
                        PERSON [MINUS] ITEM MEASURE
```

Figure 2: Expected score ICC: means.

Cross validation

In order to check the accuracy of the link, a small sample of students at different locations along the ability scale can be selected. It is better to select students whose ability measures on the test (after being equated with the rater-based analysis) fall well in the middle of the bands and students who fall very close to the transition points. Then the group of expert raters who rated the items can interview these students and try to rate them on the proficiency scale, they rated the items on. Agreements between rater judgments of where the students fall on the proficiency scale and students' measures, which empirically put them at certain levels on the scale, confirm the equating. Disagreements can be examined in case they indicate the need for slight adjustments to the criterion levels thresholds.

Purya Baghaei

Linking Terminology: Raw Score and Rasch

Term	Linn & Mislevy meaning	Rasch meaning
Linking	general term for making the results of different tests comparable	enabling the data to be analyzed together in one analysis (if desired) to construct one overall set of measures
Equating	correspondence of raw scores between tests	putting the measures in the same frame of reference
Calibration	putting the scores in the same frame of reference	constructing item measures in the internal frame of reference
Projection	scores on one test weakly predict scores on another test	(a height-weight situation)
Moderation	equivalences based on matching up sample statistics	(capitalizing on accidents in the data)
Anchoring (fixing)	-	measures obtained from one analysis (or construct theory) imposed on another to place it in the same frame of reference.
Local origin	zero score or sample mean	reference location from which to measure along the latent variable

Linn, R. L. (1993). Linking results of distinct assessments. Applied Measurement in Education, 6, 83-102.

Mislevy, R. J. (1992). Linking educational assessments: Concepts, issues, and prospects. Princeton, NJ: Educational Testing Service.

Rasch SIG Reception

The Rasch SIG is planning to have a reception for SIG members during the 2007 AERA conference in Chicago. It is schedule for 8:00pm on Tuesday April 10, 2007. The reception is being sponsored by Pearson-VUE. It includes a buffet dinner. The bar will be cash. As a convenience, the reception will be held at a near by hotel which is easily within walking distance from the Hyatt. There is limited seating so reserve a seat as soon as you know that you can make it.

Please RSVP to Matt Bennett mbennett@ncsbn.org and include:
1. Your name
2. Number of seats (1 or 2)
3. Your contact information (email and phone)

Edward W. Wolfe, Secretary, Rasch Measurement SIG

"Managers who don't know how to measure what they want, settle for wanting what they can measure."
f-Law no. 51 of Russell Lincoln Ackoff.
Management f-Laws; Russell L Ackoff and Herbert J Addison, with considered responses by Sally Bibb; Triarchy Press

MetaMetrics Workshop Series on Psychometrics

An Introduction to Rasch Measurement: Theory and Application

by David Andrich

March 26-29, 2007 - Monday-Friday
Durham, North Carolina

You are invited to a free four-day workshop introducing the theory and applications of Rasch measurement and providing hands-on experience with RUMM2020 data analysis software. The workshop will combine lecture, question-and-answer and small-group instruction. You will have opportunities to analyze your own data.

We will study principles of the Rasch models from the perspective of the Item Characteristic Curve and Differential Item Functioning. RUMM2020 will be used to demonstrate concepts and teach you how to analyze data in a flexible way. Case studies will be used that bring together the professional understanding of the variable of assessment, item construction, and the use of statistical indices in determining the validity of item sets.

Instructional material will apply Rasch models to dichotomous (multiple choice) and polytomous (rating scale and partial credit) data. Familiarity with Microsoft Excel, basic statistics and the Windows platform is a plus.

More information and registration details at:
http://www.lexile.com/DesktopDefault.aspx?view=re&tabindex=3&tabid=92

David Andrich

Professor David Andrich, has been named the inaugural Chapple Chair in Education at The University of Western Australia, commencing March 1, 2007. Vice-Chancellor Professor Alan Robson said the appointment marked a significant occasion in the university's history: "Professor Andrich is a leading international figure in theoretical studies of educational measurement and assessment and he is well known in Western Australia for his reports to the Curriculum Council on tertiary entrance assessment. His appointment is in keeping with the university's commitment to achieving international excellence."

"Measure twice, cut once" - the Golden Rule of Carpentry

Rasch Workshop

Hands-on Introduction to IRT/Rasch Measurement Using Winsteps

by Ken Conrad & Barth Riley

March 26-27, 2007 - Monday-Tuesday
University of Illinois - Chicago

Social scientists have great need for the development of valid measures, e.g., of the quantity and quality of health services and of the outcomes of those services. Many researchers are frustrated when existing instruments are not well tailored to the task, since they then cannot expect sensitive, accurate, or valid findings. This workshop presents the theory and practice of classical test theory, the traditional approach. It then provides an overview of modern measurement as practiced using item response theory with a focus on Rasch measurement. Rasch analysis provides the social sciences with the kind of measurement that characterizes the natural sciences. Since Rasch focuses on the items and the persons rather than the test score, the synthesis of quantitative analysis with qualitative issues is experienced in a way that is rare in social science. Ultimately, Rasch measurement can facilitate more efficient, reliable, and valid assessment while improving privacy and convenience to users. The Workshop is useful for anyone who wants to understand the role of modern measurement in research.

Attendees will learn hands-on:
* Differences between Classical Test Theory and Rasch
* Why and how Rasch creates linear, interval measures
* The inner workings of the Rasch model
* How to run Winsteps analyses
* Interpretation of Rasch/Winsteps output

You need a recent Winsteps running on a lap-top computer. We provide Winsteps free, but time-limited.

For more details and registration:
www.winsteps.com/workshop.htm

AERA Annual Meeting, Chicago, 2007: Rasch-related Activities

Sunday, April 8, 2007

2.010 A Hands-On Introduction to Latent Class Models, Mixture Rasch Models, and Diagnostic Mixture.

Sun., Apr 8, 9:00-5:00. Fairmont, Regent Room, Third Level.

Professional Development Training. *Matthias von Davier (ETS)*

Monday, April 9, 2007

16.074 Educational Measurement Applications I. SIG-Rasch Measurement.

Mon, Apr 9, 12:00-2:00. Hyatt / Field, West Tower-Silver Level.

Chair: *Gene A. Kramer (American Dental Association)*

A Study of Confidence and Accuracy Using the Rasch Modeling Procedures. *Insu Paek (ETS), Lazar Stankov (ETS), Jihyun Lee (ETS), Mark R. Wilson (University of California-Berkeley)*

A Comparison of Traditional and IRT Scoring Rules for Time-Limit Tests. *Margo G.H. Jansen (University of Groningen), Margaretha P.C. van der Werf (RION Institute for Educational Research), Hans Kuyper (University of Groningen)*

Comparing Parameter Recovery Accuracy Between the Rasch Testlet Model and the One-Parameter Multilevel Testlet Model. *Wei He (Michigan State University), Hong Jiao (Harcourt Assessment, Inc.), Shudong Wang (Harcourt Assessment, Inc.), Chueh-An Hsieh (Michigan State University)*

Optimizing Item Pool Characteristics to Control Item Exposure in a Computerized Adaptive Test. *Cherdsak Iramaneerat (University of Illinois-Chicago), John A. Stahl (Promissor, Inc.)*

Transferring IRT Scale Scores Using an Equipercentile Linking Method. *Daeryong Seo (Harcourt Assessment, Inc.), Husein M. Taherbhai (Harcourt Assessment, Inc.)*

Using the Rasch Measurement Model and the Bookmark Standard-Setting Procedure to Establish Cut-scores on the STOU-TBS Test. *Sungworn Ngudgratoke (Michigan State University), Ratchaneekool Pinyopanuwat (Sukhothai Thammathirat Open University), Nalinee Na Nakorn (Sukhothai Thammathirat Open University)*

Discussant: *Jon S. Twing (Pearson Educational Measurement)*

16.092 Measurement Issues. SIG-Rasch Measurement

Mon, Apr 9, 12:00-12:40. Hyatt / Grand Ballroom, Sections E-F, East Tower-Gold Level

Stability of Rasch Scales Over Time. *Catherine S. Taylor (University of Washington), Yoonsun Lee (Office of Superintendent of Public Instruction)*

A Phenomenology of Quantity for Social Science Applications. *William P. Fisher (Avatar, International, Inc.)*

Thinking Mathematically: *Eavesdropping on the Complicated Conversations in Probabilistic Models. Sharon G. Solloway (Bloomsburg University)*

16.110 School Community, Climate, and Culture.

Mon, Apr 9, 12:00-12:40. Hyatt / Grand Ballroom, Sections C-D South, East Tower-Gold Level

Characteristics of Successful Schools: A School Climate Survey. *Stacie Ann Hudgens (Learning Point Associates), Everett V. Smith (University of Illinois-Chicago)*

17.019 Methodological Issues in Survey Research as Applied in Educational Settings.

Mon, Apr 9, 12:50-1:30. Hyatt / Grand Ballroom, Sections C-D North, East Tower-Gold Level

Rasch Workshop

An Introduction to Rasch Measurement: Theory and Applications

by Everett V. Smith Jr. & Richard M. Smith

April 7-8, 2007 - Saturday-Sunday immediately before AERA

University of Illinois - Chicago

This training session on the theory and applications of Rasch measurement will provide participants with the necessary tools to become effective consumers of research employing Rasch measurement and the skills necessary to solve practical measurement problems. Instructional material will be based on four Rasch measurement models: dichotomous, rating scale, partial credit, and many-facet data. Participants will have the opportunity to use current Rasch software.

The format will consist of eight units:

- Introduction to Rasch Measurement
- Item and Person Calibration
- Dichotomous and Polytomous Data
- Performance and Judged Data
- Applications of Rasch Measurement I and II
- Examples of Rasch Analyses
- Analysis of Participants' Data.

The material covered is these units is an overview of material that would normally be covered in approximately three graduate level measurement courses. Registration includes the full 2-day workshop, a continental breakfast each morning, over 550 pages of handouts and tutorial material, a copy of *Introduction to Rasch Measurement* (a 698 page book) and a one-year subscription to the *Journal of Applied Measurement.*

For more details and registration: www.jampress.org under Rasch Measurement Workshops

Enhanced Reporting of Survey Data: A Psychometric Approach. *Andrew Swanlund (Learning Point Associates), Stacie Ann Hudgens (Learning Point Associates), Chloe R. Hutchinson (Learning Point Associates)*

Measuring Individual Preferences for Counselor Characteristics. *Jennifer Ann Weber (University of Kentucky), Kelly D. Bradley (University of Kentucky)*

19.099 Educational Measurement Applications II. SIG-Rasch Measurement

Mon, Apr 9, 2:15-2:55. Hyatt / Grand Ballroom, Sections E-F, East Tower-Gold Level

Analysis of Study Skills: Self-Efficacy of Hong Kong High School Students. *Qiong Fu (University of Illinois-Chicago), Man-Tak Yuen (University of Hong Kong), Everett V. Smith (University of Illinois-Chicago)*

Defining the Profession: A Job Task Analysis for the Federation of State Massage Therapy Boards. *Donna J. Surges Tatum (Meaningful Measurement, Inc.), Johnna Gueorguieva (University of Illinois-Chicago)*

Development of a Diagnostic Reading Assessment Battery Using Rasch Measurement. *Kim H. Koh (Nanyang Technological University), Susan Bee-yen Gwee (National Institute of Education-Singapore)*

20.024 Studies of Vocabulary Instruction and Acquisition.

Mon, Apr 9, 3:05-3:45. Hyatt / Grand Ballroom, Sections E-F, East Tower-Gold Level

Using Rasch Measurements to Analyze the Difficulty of Target Words and Preschoolers' Vocabulary Ability From Read-Alouds. *Cynthia B. Leung (University of South Florida-St. Petersburg)*

24.050 Division D: New Member Poster Session.

Mon, Apr 9, 6:15-7:45. Hyatt / Riverside Center Exhibition Hall, East Tower-Purple Level

Effects of Linking Design, Growth Pattern, and Dimensionality on Vertical Scaling. *Shudong Wang (Harcourt Assessment, Inc.), Hong Jiao (Harcourt Assessment, Inc.), Michael J. Young (Harcourt Educational Measurement)*

Using the Monte Carlo Procedure to Test the Significance of Local Item Dependence. *Wei Tao (Boston College)*

Separate Versus Concurrent Calibration Methods With Different Estimation Methods for Vertical Scaling With the Rasch Model. *Shu-Ren Chang (Rockford Public Schools), Che-Ming A. Lau (Harcourt Assessment, Inc.), Shu-Mei Lien (University of Nebraska-Lincoln), Yue Zhao (University of Massachusetts-Amherst), Wendy Lam (Harcourt Assessment, Inc.)*

Gender Differences and Similarities in PISA 2003 Mathematics: A Comparison Between the United States and Hong Kong. *Ou Lydia Liu (University of California-Berkeley), Mark R. Wilson (University of California-Berkeley)*

Tuesday, April 10, 2007

29.102 Cognitive, Social, and Motivational Processes: Paper Discussion (Session 3)

Tue, Apr 10, 8:15-8:55. Hyatt / Grand Ballroom, Sections C-D North, East Tower-Gold Level

Are Immigrant Students More Motivated in Mathematics? The Effect of Response Tendencies in PISA 2003. *Päivi Taskinen (IPN), Karin Zimmer (IPN), Steffen J. Brandt (Leibniz – Institut für die Pädagogik der Naturwissenschaften, Universität Kiel)*

31.011 Challenging Times for Adolescents: Insights From Large-Scale Datasets From Around the World.

Tue, Apr 10, 10:35-12:05. Sheraton / Chicago Ballroom, Section X, Level 4

Teacher and Principal Perspectives on Student Victimization and School Connectedness in Israel: A Rasch Analysis of Multiple Views of Violence in the Same Schools. *Susan I. Stone (University of California-Berkeley), Ron Avi Astor (University of Southern California), Rami Benbenishty (Hebrew University of Jerusalem)*

31.088 Advances in Measurement Theory and Method. SIG-Rasch Measurement

Tue, Apr 10, 10:35-12:05. Hyatt / Addams, West Tower-Silver Level

Chair: *Karen L. Draney (University of California-Berkeley)*

The Construct Underlying Seven Aberrance Indices. *Jing Chen (American Institutes for Research), Rui Gao (ETS), Ying Lu (ETS)*

Direct and Indirect Year-to-Year Linking Design in Mixed-Item Format Test Under the Rasch/Partial Credit Model. *Daeryong Seo (Harcourt Assessment, Inc.), Husein M. Taherbhai (Harcourt Assessment, Inc.), Che-Ming A. Lau (Harcourt Assessment, Inc.), Timothy P. O'Neil (University of Massachusetts)*

Investigating Displacement in the Rasch Model. *John A. Stahl (Promissor, Inc.), Timothy Joseph Muckle (Pearson VUE), Betty A. Bergstrom (Promissor, Inc.), James S. Masters (University of North Carolina-Greensboro), Kirk A. Becker (Pearson VUE)*

Evaluating the Accuracy of Item Parameter Estimates and Standard Error of Estimates That WINSTEPS Reports. *Wei He (Michigan State University), Mark D. Reckase (Michigan State University)*

Using the Rasch Model to Confirm the Effectiveness of Rating Scale Categorizations. *Nicholas D. Myers (University of Miami), Deborah L. Feltz (Michigan State University), Edward W. Wolfe (Virginia Polytechnic Institute and State University)*

Discussant: *Martha S. McCall (Northwest Evaluation Association)*

34.025 Research on Equating.

Tue, Apr 10, 12:25-1:55. Marriott / Cook, Third Floor

A Comparison of IRT Equating Methods on Recovering Parameters and Capturing Growth in Mixed-Format Tests. *Su G. Baldwin (University of Massachusetts-Amherst), Peter Baldwin (University of Massachusetts), Michael L. Nering (Measured Progress)*

Investigating Scaling Effects on Alternate Test Forms Using the Common Item Approach in Rasch-Based Equating. *Nathan L. Wall (Harcourt Assessment, Inc.), Qing Yi (Harcourt Assessment, Inc.)*

37.085 Measurement Applications in Early Education, Special Education, and Early Intervention

Tue, Apr 10, 2:15-3:45. Hyatt / Atlanta, East Tower-Gold Level

Chair: *Seock-Ho Kim (University of Georgia)*

Applying the Rasch Model to Guide Refinement of Early Childhood Individualized Family Service Plans. *Lee Ann Jung (University of Kentucky), Kelly D. Bradley (University of Kentucky), Shannon O. Sampson (University of Kentucky)*

Which Standardized Measure of Classroom Quality Is Valid: ECERS or ELLCO? *Eilene Edejer (Chicago Public Schools), Nikolaus Bezruczko (Institute for Objective Measurement, Inc.)*

Using the Rasch Model to Determine Equivalence of Forms in the Trilingual Lollipop Readiness Test. *William S. Lang (University of South Florida), Judy R. Wilkerson (Florida Gulf Coast University)*

Expanding the Concept of Educational Quality: Parents' Perceptions of Special Education and Early Intervention Services. *William P. Fisher (Avatar, International, Inc.), Batya Elbaum (University of Miami), Alan Coulter (Louisiana State University), Lisa Persinger (Louisiana State University)*

Discussant: *Edward W. Wolfe (Virginia Polytechnic Institute and State University)*

37.097 Psychometrics

Tue, Apr 10, 2:15-2:55. Hyatt / Grand Ballroom, Sections C-D South, East Tower-Gold Level

Interval Estimation for Proficiency in Multiple-Format Testing. *Chiou-Yueh Shyu (National University of Taiwan)*

Taught in a Foreign Language, Tested in the Mother Tongue: Advantage or Disadvantage for Test-Takers? *Markus Broer (Supreme Education Council of Qatar), Juan E. Froemel (Evaluation Institute. Supreme Education Council), Richard Schwarz (CTB/McGraw-Hill)*

44.043 Rasch Measurement-SIG Business Meeting Unit: SIG-Rasch Measurement

Tue, Apr 10-6:15-8:15 Hyatt, Atlanta, East Tower-Gold Level

Invited speaker: *George Engelhard, Jr.*

Wednesday, April 11, 2007

50.024 Research in Early Childhood Mathematics Education

Wed, Apr 11, 10:35-12:05. Hyatt / Addams, West Tower-Silver Level

Development of a Measure of Early Mathematics Achievement Using the Rasch Model. *Douglas H. Clements (State University of New York-Buffalo), Julie Sarama (State University of New York-Buffalo), Xiufeng Liu (State University of New York-Buffalo)*

50.028 Issues in Estimating Model Parameters

Wed, Apr 11, 10:35-12:05. Marriott / Dupage, Third Floor

Recovery of Parametric Item Characteristic Curves With Parametric and Nonparametric IRT Models. *Qiong Wu (Pennsylvania State University), Pui-Wa Lei (Pennsylvania State University)*

50.081 Item Wording and Order Effects in Survey Research

Wed, Apr 11, 10:35-12:05. Marriott / Chicago Ballroom, Section D-Fifth Floor

A Transverse Study of Items' Wording Impact With Rasch's Rating Scale Model. *Jean-Guy Blais (University of Montréal), Julie Grondin (University of Montréal), Nathalie Loye (University of Ottawa), Gilles Raiche (Université du Québec-Montréal)*

51.073 Technical Issues in Large-Scale Assessment

Wed, Apr 11, 12:25-1:55. Sheraton / Michigan, Level 2

Evaluating Different Person-Fit Indices to Detect Inappropriate Cases in a Rasch Model Calibration Context. *Seon-Hi Shin (Harcourt Assessment, Inc.), Yoonsun Lee (Office of Superintendent of Public Instruction), Se-Kang Kim (Fordham University)*

On the Estimation of Classification Consistency Indexes for Complex Assessments. *Matthew Stearns (Data Recognition Corporation), Richard Smith (Journal of Applied Measurement)*

51.079 Measuring Educational Quality. SIG-Rasch Measurement

Wed, Apr 11, 12:25-1:55. Hyatt / Plaza Ballroom, Section A, East Tower-Green Level

Chair: *G. Gage Kingsbury (Northwest Evaluation Association)*

Measurement of Master's Degree Thesis Quality on a Linear Scale. *A. A. Maslak (Slavyansk-on-Kuban State Pedagogical Institute), Tatyana S. Anisimova (Slavyansk-on-Kuban State Pedagogical Institute), Nikolaus Bezruczko (Institute for Objective Measurement, Inc.)*

Rasch Model Validation and Application of a Linear Scale of Teacher Observations of School Principal Leadership. *Robert Frederick Cavanagh (Curtin University of Technology), Graham B. Dellar (Curtin University of Technology), Joseph Thomas Romanoski (Curtin University of Technology)*

An Examination of the Psychometric Properties of the Graduate Student Advising Survey. *Benita J. Barnes (University of Massachusetts-Amherst), Linda A. Chard (Michigan State University), Edward W. Wolfe (Virginia Polytechnic Institute and State University)*

Why Did Teachers Stay, Move, or Leave? A Practical Application of the Rasch Model in Teacher Professional Satisfaction. *Yun Xiang (Boston College)*

What College Students Think About Their Science Teachers: A Rasch Analysis. *Maria Azucena Balberan Lubrica (Benguet State University), Joel Vizconde Lubrica (Benguet State University)*

Discussant: *Kathy E. Green (University of Denver)*

51.080 Multifaceted Measurement of Judged Performances. SIG-Rasch Measurement

Wed, Apr 11, 12:25-1:55. Hyatt / Burnham, West Tower-Silver Level

Chair: *Ronald T. Mead (DRC)*

An Application of the Multi-facets Rasch Model Analysis and Factor Analysis for Oral Examinations. *Surintorn Suanthong (Measurement Research Associates, Inc.), Mary E. Lunz (Measurement Research Associates, Inc.)*

Considerations in Developing a Benchmark Scale for Many-Facet Rasch Analysis. *Ross M Brown (Measurement Research Associates, Inc.), Mary E. Lunz (Measurement Research Associates, Inc.)*

Many-Facet Rasch Analysis of Student Evaluation. *Zongmin Kang (University of Toledo), Gregory E. Stone (University of Toledo)*

A Multifacet Rasch Analysis of the Teacher Candidate Disposition Assessment. *Susan M Gracia (Rhode Island College)*

Discussant: *Steven Stemler (Wesleyan University)*

53.023 Research in Physical Science Education

Wed, Apr 11, 2:15-3:45. Hyatt / DuSable, West Tower-Silver Level

The ChemQuery Story: Measurable Insights on How Students Learn Chemistry. *Jennifer M. Claesgens , Kathleen Scalise , Karen L. Draney , Mark R. Wilson , Angelica Stacy (University of California-Berkeley)*

Thursday, April 12, 2007

58.041 Investigating the "Knowledge of Reading" Needed to Teach Elementary Students to Read: The Role of Conceptualization, Measurement, and Evidence

Thu, Apr 12, 8:15-10:15. Hyatt / Columbus Hall, Section C, East Tower-Gold Level

Connecting Primary Grade Teacher Knowledge to Primary Grade Student Achievement: Developing an Evidence-Based Reading/Writing Teacher Knowledge Assessment System. *D. Ray Reutzel (Utah State University), Janice A. Dole (University of Utah), Parker C. Fawson (University of Utah), Sylvia Read (Utah State University), Richard R. Sudweeks (Brigham Young University)*

60.077 New Developments in Measurement Thinking. SIG-Rasch Measurement

Thu, Apr 12, 10:35-12:05. Hyatt / Burnham, West Tower-Silver Level

Chair: *Chien-Lin Yang (American Dental Association)*

Standards for Objective Tests. *Agustin Tristan-Lopez (IEESA)*

Dialogue Between Measurement and Practice in Rating Scale Structure Analysis. *Joel Vizconde Lubrica (Benguet State University)*

Linear Model to Assess the Scale's Validity of a Test. *Agustin Tristan-Lopez (IEESA)*

Validation of Students' Learning-Strategy Scale Using a Multidimensional Rasch Measurement Model. *Daeryong Seo (Harcourt Assessment, Inc.), Husein M. Taherbhai (Harcourt Assessment, Inc.), Yong-Hwi Park (KDE)*

Discussant: *Jon S. Twing (Pearson Educational Measurement)*

61.108 Science Teaching and Learning (STL-SIG) Poster Session

Thu, Apr 12, 12:25-1:55. Hyatt / Riverside Center Exhibition Hall, East Tower-Purple Level

Evaluating and Restructuring Science Assessments: An Example Measuring Students' Conceptual Understanding of Heat. *Kelly D. Bradley (University of Kentucky), Jessica Dawn Cunningham (University of Kentucky), Shannon O. Sampson (University of Kentucky)*

66.012 Emerging Scholars and Scholarship in Education Research: AERA, NAEd, and IES Postdoctoral Fellows and Their Work

Thu, Apr 12, 4:05-6:05. Sheraton / Sheraton Ballroom, Section I, Level 4

Creating a Metric for Measuring Early Student Literacy Development: A Rasch Analysis of DIBELS Assessment Data. *Gina Biancarosa (Stanford University), David W. Kerbow (University of Chicago), Anthony S. Bryk (Stanford University)*

Friday, April 13, 2007

72.062 Multilevel Measurement Models and Issues

Fri, Apr 13-8:15-10:15 Marriott / Chicago Ballroom, Section G-Fifth Floor

A Comparison of Three DIF Detection Procedures Using Hierarchical Generalized Linear and Nonlinear Mixed Models. *Yuk F. Cheong (Emory University), Akihito Kamata (Florida State University)*

74.024 Cognitive Analysis and Dimensionality

Fri, Apr 13, 10:35-12:05. Marriott / Chicago Ballroom, Section H-Fifth Floor

Applications of a Rasch Model With Subdimensions. *Steffen J. Brandt (Leibniz – Institut für die Pädagogik der Naturwissenschaften, Universität Kiel)*

74.039 Noncognitive Assessments

Fri, Apr 13, 10:35-12:05. Marriott / Chicago Ballroom, Section F-Fifth Floor

The Impact of Candidate Communication Ability on Candidate Oral Examination Performance. *Mary E. Lunz (Measurement Research Associates, Inc.), Philip G. Bashook (University of Illinois)*

DIF matters:

A practical approach to test if Differential Item Functioning makes a difference

Differential Item Functioning (DIF) in psychometric tests has long been recognized as a potential source of bias in person measurement. Originally called `item bias' (Lord, 1980), the analysis of DIF is concerned with identifying significant differences, across group membership, of the proportion of individuals *at the same apparent ability level* who answer a given item correctly (or can do a particular task). If an item measures the same ability in the same way across groups then, except for random variations, the same success rate should be found irrespective of the nature of the group. Items that give different success rates for two or more groups, at the same ability level, are said to display DIF (Holland & Wainer, 1993). When developing new tests, items displaying DIF would normally be revised or discarded.

Existing tests may also contain items displaying DIF. Sometimes summary and individual item fit statistics are satisfactory, yet DIF is still apparent. If DIF occurs within a Rasch model framework, it may be productive to treat items exhibiting DIF as different items for different groups. This process is called "splitting for DIF". It produces DIF-free person estimates (Tennant A, et al, 2004), but the data manipulation can be complex and time-consuming (Hambleton 2006).

Another issue is that of "cancellation of DIF" (Teresi JA, 2006). This is where some items favor one group and other items the other group. In practice, DIF always balances out as it is conditional upon the raw score. That is, at a given level of the trait corresponding to an overall score of 'x', we would expect members of a group to have a particular success rate on an item. When this success rate is considerably less due to DIF against the group, then the group's overall success-rate needs to be made up from elsewhere in order for the group members obtain the score of 'x'. The success-rate may be made up from another item which balances the original DIFed item, or the counter-balancing effect may be spread across many items.

Our Study

Although two types of DIF can be identified – uniform DIF (constant across ability levels) and non-uniform DIF (varying across ability levels. In our simulation study, we looked only at Uniform DIF. To do this, we simulated four datasets:

SET A: a 20 item 5 category (0-4) response set with 400 cases, divided evenly between males and females, where two items are simulated to have DIF, both giving males a higher expected score on each item.

SET B: replicated the first set except with 10 items.

SET C: replicated the first set except with just 5 items.

SET D: replicated SET B but with DIF adjusted so that one item favored males, the other item females, both by one logit, i.e., perfect cancellation.

All datasets were simulated to have the same item difficulty range and the same rating scale structure. Item difficulty was modified just for the two items showing DIF in the relevant part of the sample, to represent DIF by gender. These different sets would be typical of many scales (or subscales) used in medical outcome studies.

Approach

Items were first fitted to the Rasch-Masters Partial Credit model with the RUMM2020 program. In our study, DIF is examined through response residuals. When person *n* encounters item *i,* the observed response is X_{ni} and the corresponding expected response is $E[X_{ni}]$, with model variance $V[X_{ni}]$. The standardized residual Zni is given by

$$z_{ni} = \frac{X_{ni} - E[X_{ni}]}{\sqrt{V[X_{ni}]}} \ . \qquad (1)$$

Then each person is assigned to a factor group (e.g., gender) and classified by ability measure on the latent trait into one of G class intervals. Then, for each item, the observation residuals are analyzed with a two-way analysis of variance (ANOVA) or person factor by class interval. The presence of DIF is indicated by statistically significant inter-person-group variance.

Once DIF was identified using ANOVA, the strategy we adopted is a variation of the iterative 'top-down purification' approach, in which the requirement for assessing DIF is a baseline set of 'pure' items (Lord, 1980). In our approach we identified the 'pure' item set by removing items displaying DIF. Given the pure set, the item parameter estimates for the three items displaying the least DIF (2 items for the Set C 5-item simulation) were exported to an anchor file. The original full set of items was then re-run anchored by those three items so that person estimates were based upon the measurement framework defined by the anchored items which show the minimum DIF. This accords with the measurement framework of the pure analysis. The person estimates from the two analyses (pure and full-with-anchors), plus the standard errors of the estimates, were then exported into Excel and compared.

Irrespective of the amount of DIF detected, we argue that for practical purposes, given satisfactory fit to the model, if the person estimates remain largely unchanged, then the DIF is trivial and can be ignored. We define a trivial impact as being a difference in the person estimates from the two analyzes of less than 0.5 logits (Wright & Panchapakesan, 1969).

A final analysis confirms the unidimensionality of the full data set, to make sure this has not been compromised by DIF. A principal-components analysis of the residuals identifies positive and negative loading subsets of items which are used to generate estimates for comparison using a series of independent t-tests. Where the number of significant individual t-tests does not exceed 5% (or the lower bound of the binomial confidence interval does not) then the scale is unidimensional (Smith, 2002).

Routine summary fit statistics are reported for each simulation, including item-and person mean residuals and their standard deviations, which would be zero and one respectively under perfect fit. A Chi square interaction value reports on invariance across the trait, and would be non-significant where data meet model expectations. A Person Separation Index is also reported, equivalent to Cronbach's Alpha, as a measure of person sample "test" reliability.

Our sample size of 200 per group is sufficient to test for DIF in the residuals where at α of p<0.05, β of p<0.20 the effect size between groups is 0.281. Bonferroni corrections are applied to both fit and DIF statistics due to the number of tests undertaken for any given scale (Bland & Altman, 1995).

Our Findings

In the analyses of all 4 datasets, the two items simulated to have DIF were reported to have significant DIF, only one other item, in SET C, was reported to have significant DIF. All 4 datasets showed good fit to the Rasch model, and were not rejected by the test of unidimensionality at the 5% level. The crucial results are shown here:

SET	Total items	DIF items	Person measure differ >.05 logits	Other findings
A	20	2 M+	0.75%	
B	10	2 M+	4%	
C	5	2 M+	39.4%	1 item with compensatory DIF
D	10	1 M+ 1 F+	0.00%	
M+ favors Males			Sample: 400 (200 M, 200 F)	

In each simulated dataset, the ANOVA-based DIF analysis detected the simulated uniform-DIF items. Only in SET C did an additional DIF item emerge and this was purely compensatory to the simulated DIF. Although not significant, all items showed some level of DIF, indicating how the presence of two items favoring males forced other items to favor females. This 'cancellation' phenomenon is well known and raises some important issues (Teresi JA, 2006). It has been argued removal of the items with the most severe level of DIF may actually induce more, rather than less, DIF at the test score level.. Thus total test scores can meaningfully be used for the comparison of populations, even though they are made up of items that are individually DIFed (Borsboom,2006).

An example DIF of behavior (for Item 1 of SET B) is shown in Figure 1. This pictures how uniform DIF puts one group to the left (easier) side of the Rasch ICC, and the other group to the right (harder) side.

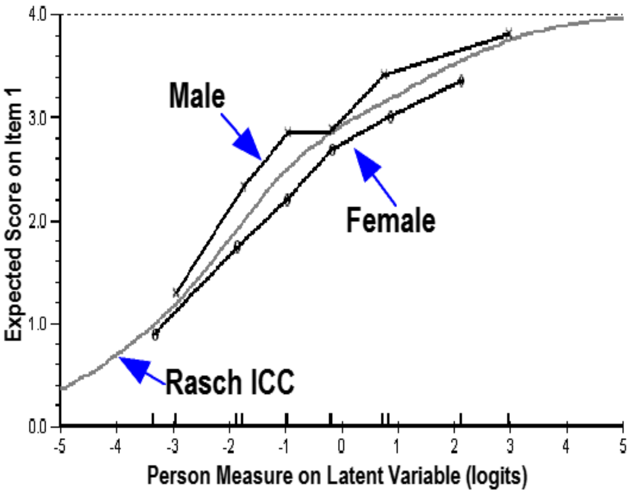


Figure 1. DIF on item 1 of the 10 item set (SET B).

SET C raises the DIF content to 40% by reducing the item set to five items. In this dataset, one of the items simulated without DIF demonstrated compensatory DIF, although this disappeared when the two simulated DIF items were removed to obtain the pure set. Compensatory DIF is where one or more items are forced to compensate for primary DIF, as in our two simulated DIF items. However, the extent of cancellation (does it fully compensate?) remains an empirical question. Thus, while all fit statistics showed fit to model expectation, on this occasion, 39.4% of the sample recorded a difference in estimates greater than 0.5 logit.

In SET C, where person estimates differ considerably, the mean difference between males and females increased from 0.18 logits to 0.33 logits in the unDIFed and DIFed item sets respectively. However, the compensatory DIF may lead the analyst to presume that DIF is canceling out, but clearly there is a significant impact on individual estimates, and some impact on group estimates. Thus any analysis where gender (or any other factor) is a substantive component may detect gender differences where they do not exist, simply because of the DIFing effect at the test level (although in this case the difference was non significant). It seems imperative, in the presence of DIF, even with fit to model expectations, to explore the impact on person estimates and see if the DIF makes a difference.

Finally Set D, which included two exactly-canceling items, also showed good fit to model expectations, and both simulated DIF items showed up as significant. The main difference in the results for this set was that the magnitude of difference between estimates was virtually zero for all cases with no cases outside a difference of 0.5 logits (the highest difference was 0.06 logits). In SET B, which has the same number of items, but where two are DIFed in the same direction, 4% of person estimates differed by 0.5 logits or more, so here the DIF is not fully compensated.

2ª Reunión Regional Norte, Centro América y Caribe de Evaluación Educativa

Español Spanish Language Meeting

Campeche, Mexico, 26-28 September 2007

sponsored by the

Institute of Evaluation and Advanced Engineering

Advancements, Proposals and Solutions

1. Test design and measurement
2. Test process, administration and certification
3. Reports and diffusion of test results

The idea of the meeting is that participants present their work and their problems, **in order to get a solution** or, at least, an idea to improve their tests, from the other participants and the expert panelists.

Conference presentations by leading researchers from Spain, Colombia, Honduras and Mexico.

Details at: www.rasch.org/regional.htm

Further information from: marcelt_84@yahoo.com

Agustin Tristan

The results here suggest that the level of substantive DIF may make a difference at both the individual and group level and thus needs to be recognized and routinely reported. This simple strategy, of comparing the estimates from the full set of items with that from the 'pure' set (the former anchored to the most pure items of the latter) is one way of detecting the impact of DIF.

This simple simulation study has shown that it is possible to examine the impact of DIF in existing scales by looking at the effect upon person estimates. This has shown that these estimates may differ considerably under certain conditions, generally linked to the proportion of DIFed items in the test. While exact cancellation results in no difference for person estimates, compensatory DIF may not fully cancel the DIF, even when significant DIF items emerge as a result.

These findings have important implications particularly for those scales used in routine clinical practice. Some scales have clinical cut points which suggest the need for treatment, and little DIF can be tolerated under these circumstances. Further evidence also needs to be sought for the effect of DIF at the group level as increasingly, in international clinical trials, data are pooled from different countries, and so the scales must be invariant by culture.

Alan Tennant, University of Leeds, UK
Julie F. Pallant, Swinburne University of Technology, Australia

Bland JM, Altman DG. (1995). Multiple significance tests: the Bonferroni method. *British Medical Journal, 310, 170.*

Borsboom D (2006) When Does Measurement Invariance Matter? *Medical Care, 44:(11) Suppl 3*

Hambleton RK (2006) Good Practices for Identifying Differential Item Functioning. *Medical Care 44: (11) Suppl 3.*

Holland PW, Wainer H (1993) *Differential Item Functioning*. Hillsdale. NJ: Lawrence Erlbaum

Lord FM (1980) *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale NJ: Lawrence Erlbaum Assoc.

Tennant A, Penta M, et al. (2004) Assessing and adjusting for cross cultural validity of impairment and activity limitation scales through Differential Item Functioning within the framework of the Rasch model: the Pro-ESOR project. *Medical Care, 42: 37-48*

Smith EV (2002) Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3:205-231.*

Teresi JA (2006) Different Approaches to Differential Item Functioning in Health Applications: Advantages, Disadvantages and Some Neglected Topics. *Medical Care, 44:S152–S170*

Wright, BD, & Panchapakesan, N (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, (29), 23-48.*

Journal of Applied Measurement
Volume 8, Number 1. Spring 2007

Attitudes, Order and Quantity: Deterministic and Direct Probabilistic Tests of Unidimensional Unfolding. *Andrew Kyngdon and Ben Richards. 1-35.*

Conception and Construction of a Rasch-Scaled Measure for Self-Confidence in One's Vocabulary Ability. *Michaela M. Wagner-Menghin. 36-47*

Relative Precision, Efficiency and Construct Validity of Different Starting and Stopping Rules for a Computerized Adaptive Test: The GAIN Substance Problem Scale. *Barth B. Riley, Kendon J. Conrad, Nikolaus Bezruczko, and Michael L. Dennis. 48-64*

Bookmark Locations and Item Response Model Selection in the Presence of Local Item Dependence. *Garry Skaggs. 65-83*

Comparing Concurrent versus Fixed Parameter equating with Common Items: Using the Rasch Dichotomous and Partial Credit Models in a Mixed Item-Format Test. *Husein M. Taherbhai and Daer Yong Seo. 84-96*

Instrument Development Tools and Activities for Measure Validation using Rasch Models: Part I – Instrument Development Tools. *Edward W. Wolfe and Everett V. Smith, Jr. 97-123*

Richard M. Smith, Editor
JAM web site: www.jampress.org

PROMS 2007

Pacific Rim Objective Measurement Symposium

July 16-19, 2007 * TaoYuan (near Taipei), Taiwan

at National College of Physical Education & Sports

You are invited to participate in this exciting event and to submit proposals for presentation. The deadline for proposal submission is **1 May 2007.** PROMS http://210.60.0.152/PROMS2007TAIWAN/ is 4 days after the International Meeting of the Psychometric Society (IMPS), Tokyo, Japan, http://www.ech.co.jp/imps2007/ .

Theme: *Objective Measurement in Diverse Disciplines*
The Conference, on Tuesday-Thursday, July 17-19, 2007, invites proposals from the introductory through advanced level on all topics related to the applications of Rasch measurement models or Item Response Theory in any disciplines, such as education, psychology, sports, languages, medicine, public health, management, sociology, and political science.

Keynote Speakers:
Dr. David Andrich, Chapple Professor, Graduate School of Education, University of Western Australia, Australia
Dr. Mark Wilson, Prof., Graduate School of Education, UC Berkeley, USA
Dr. Mike Linacre, Prof., Faculty of Health Sciences, University of Sydney, Australia
Dr. Philip E. Cheng, Research Fellow, Institute of Statistical Science, Academia Sinica, Taiwan

Invited Speakers:
Dr. Eiji Muraki, Professor, Graduate School of Educational Informatics Research Division, Tohoku University, Japan
Dr. Margaret Wu, Senior Research Fellow, Australian Council for Educational Research, Australia
Dr. Magdalena Mo Ching Mok, Professor and Centre Director, Centre for Assessment Research and Development, The Hong Kong Institute of Education, Hong Kong
Dr. Sun-Geun Baek, Professor, Department of Education, College of Education, Seoul National University, Korea

Workshops on Monday, July 16, 2007:
BILOG-MG, hosted by Dr. Eiji Muraki, chief author of BILOG-MG
ConQuest, hosted by Dr. Margaret Wu, first author of ConQuest
RUMM, hosted by Dr. David Andrich, author of RUMM
Winsteps, hosted by Dr. Mike Linacre, author of Winsteps

Organizers:
Dr. Han-Dan Yau, Professor, Graduate Institute of Sports Training Science, National College of Physical Education & Sports, Taiwan
Dr. Wen-Chung Wang, Professor, Department of Psychology, National Chung Cheng University, Taiwan

Contact:
Secretariat
Lin, Yi-Hung, Ph.D., Postdoctoral Fellow
Department of Psychology, National Chung Cheng University
168 University Rd., Min-Hsiung Chia-Yi 621, Taiwan (R.O.C.)
Phone:(886)958-353-977 * FAX:(886)5-272-0857
http://210.60.0.152/PROMS2007TAIWAN/ or
s05307@hotmail.com

Rasch-related Coming Events: 2007

March 26-27, 2007, Mon.-Tues. Introduction to IRT/Rasch Measurement Using Winsteps (Conrad & Bezruczko), Chicago www.winsteps.com/workshop.htm

March 26-29, 2007, Mon.-Thurs. MetaMetrics Workshop Series in Psychometrics – Introduction to Rasch Measurement: Theory and Application (David Andrich) (free!), Durham, North Carolina www.lexile.com

Apr. 7-8, 2007, Sat.-Sun. Introduction to Rasch Measurement: Theory and Applications, Chicago IL (Smith & Smith) www.jampress.org

Apr. 9-13, 2007, Mon.-Fri. AERA Annual Meeting, Chicago www.aera.net

May 4 - June 1, 2007, Fri.-Fri. Facets online course (Mike Linacre) www.statistics.com/courses/facets

May 2007 - Dec 2008 3-day Rasch courses, Leeds, UK http://home.btconnect.com/Psylab_at_Leeds/

June 21 - July 1, 2007, Thur.-Sun. 3rd Summer School Measurement of Latent Variables (Rasch), Russia
June 22, 2007, Fri. Workshop: Theory and Practice of Measurement of Latent Variables, Russia
www.rasch.org/russia.pdf

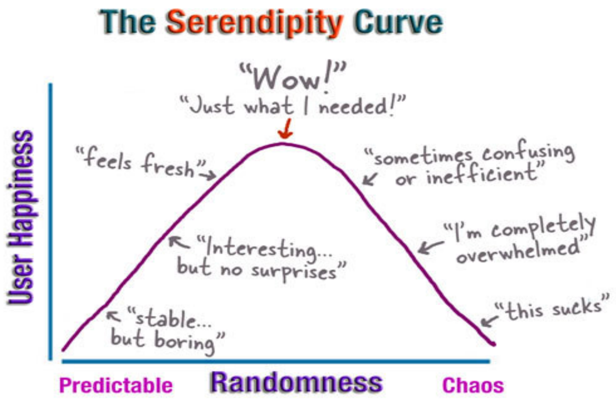
July 16, 2007, Mon.. Taiwan
ConQuest Workshop, Margaret Wu
RUMM Workshop, David Andrich
Winsteps Workshop, Mike Linacre

July 17-19, 2007, Tues.-Thurs.
Pacific Rim Objective Measurement Symposium PROMS, Taiwan
http://210.60.0.152/PROMS2007TAIWAN/

Aug. 3 - Aug. 31, 2007, Fri.-Fri. Practical Rasch Measurement with Winsteps online course (Mike Linacre) www.statistics.com/courses/rasch

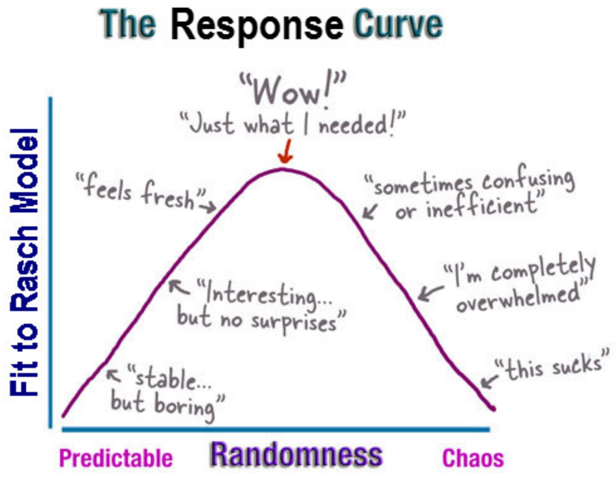
Rasch Fit and Serendipity

Kathy Sierra, author of the *Creating Passionate Users* blog at headrush.typepad.com displays this provocative Figure © 2007:



Kathy explains that, while we need predictability, we also benefit from some degree of randomness in our lives. This Figure has obvious implications for classroom management and curriculum development. Encourage an element of surprise within a stable environment.

This concept exactly matches Rasch fit statistics. Let us revise Kathy's Figure:



The essential idea behind putting many items on a test is that we learn something new from each response. That new component must neither monotonously repeat nor seriously contradict what we already know about the person and the item who interacted to generate the response.

Rasch Measurement Transactions
P.O. Box 811322, Chicago IL 60681-1322
www.rasch.org/rmt
Editor: John Michael Linacre
Copyright © 2007 Rasch Measurement SIG
Permission to copy is granted.
SIG Chair: Thomas O'Neill, Secretary: Ed Wolfe
Program Chairs: William P. Fisher & Sharon Solloway
SIG website: www.raschsig.org

Rasch Online Courses
Winsteps and Facets

May 4 - June 1, 2007, Fri.-Fri. **Facets** online course
(Mike Linacre) www.statistics.com/courses/rasch

Aug. 3 - Aug. 31, 2007, Fri.-Fri. Practical Rasch Measurement with **Winsteps** online course
(Mike Linacre) www.statistics.com/courses/facets

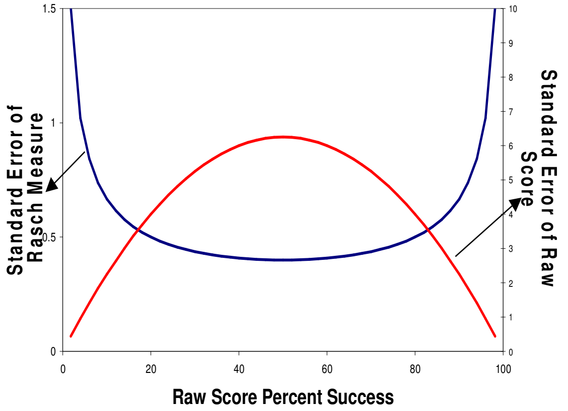
Each Course consists of 4 weeks of detailed step-by-step downloadable tutorials. There are Discussion Boards for Q-&-A and group interaction. Free time-limited versions of the software are provided. You work at your own pace.

Standard Errors and Reliabilities: Rasch and Raw Score

Question: I was taught that all raw scores on a test have the same raw score standard error, SEM, and this is:

*SEM = raw score S.D. * sqrt (1-Reliability).*

Why do standard errors for person measures differ?



Answer: The raw score "test" reliability is based on an average raw score standard error for the sample. But each raw score has a different standard error. The raw score standard errors are biggest at the center of the test and smallest at the extremes. In contrast, the standard error of a Rasch measure is smallest in the center of the test and biggest at the extremes. The plot is an idealization plot of their relationship for a 30 item dichotomous test. But, like the raw score reliability, the Rasch reliability is also based on the average standard error of the sample.

January 2008, Australia

Jan. 7-11, 2008, Mon.-Fri. Introductory course on Rasch measurement (Andrich, RUMM), Australia

Jan. 14-18, 2008, Mon.-Fri. Advanced course on Rasch measurement (Andrich, RUMM), Australia

Jan. 21, 2008, Mon. One-day RUMM Workshop (Andrich, RUMM), Australia

Jan. 22-24, 2008, Tues.-Thurs. 3rd International Conference on Measurement in Health, Education, Psychology and Marketing: Developments with Rasch models, Australia

www.rasch.org/i2008.htm