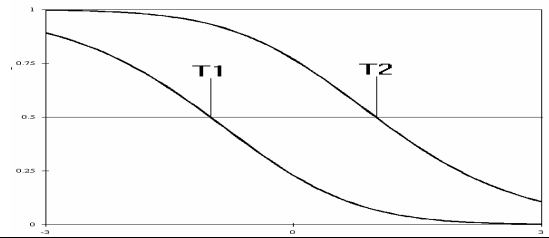
Thurstone-style Thresholds

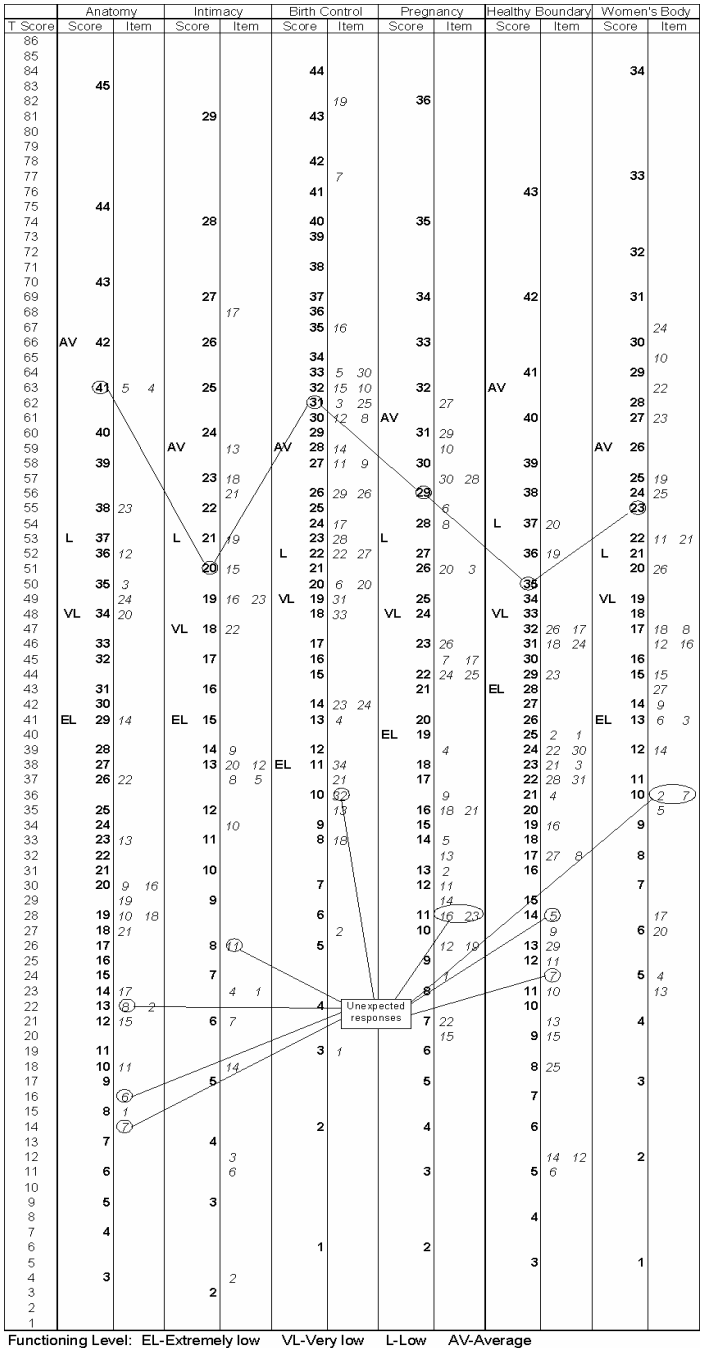


RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 16 No. 4 Spring 2003 ISSN 1051-0796

Rasch-assisted Report Form



One of the most profound benefits of Rasch measurement is the capability of producing a map of the test. The map explains the essence of the instrument. The main purpose of administering psychological tests is to assess the ability level of some trait of the individual in order to make a clinical interpretation or diagnosis. The test should be able to identify the level of functioning, strengths and weaknesses, and implications for interventions. *Knox's Cube Test-Revised* (Stone MH, Chicago, IL: Stoelting Co, 2002) is one of the few psychological instruments that provide a report form with such capability.

The Figure shows the sample report form of *Socio-Sexual Knowledge and Attitude Test-Revised* (SSKAAT-R, Griffith D & Lunsky Y, Chicago, IL: Stoelting Co, 2003). SSKAAT-R is an instrument that measure 6 domains of the sexual knowledge of individual with retardation. Each scale is Rasch analyzed separately. A T-score, a simple linear transformation of the Rasch logits scale, is given which has a mean score of 50 and a standard deviation of 10. The item number on the map indicates the difficulty of an item on T-score scale. This map provides individual item responses and ability estimates of each scale with criterion measures (the mean ability estimate of the different level of mental retardation.) Distance to ceilings and floors can also be examined as it is clear that the, Anatomy and Healthy Boundaries scale has a lower ceiling, not a critical problem in this type of a test. These interpretations can easily be made visually and intuitively from this map. The domain level knowledge can be assessed by examining overall ability and the content level knowledge can be assessed by examining the item response pattern. The clinician does not have to know the

Table of Contents

value of infit or outfit, just simply examine item response pattern and evaluate the content of the item. For example on figure 1, domain Anatomy has 3 unexpected responses in which items (6, 7, and 8) have similar contents. This is the information clinicians need to know, not the statistical parameter of the individual.

The Rasch property of population invariance, or distribution independence, greatly helps in the analysis of SSKAAT-R data since the target population has a highly skewed distribution (in terms of IQ). A fit analysis assisted to construct statistically appropriate scales. The common Rasch ruler connects items and scores. These are common knowledge in the field of measurement but definitely not for people who actually use the instruments and for those that information is probably of limited importance. The most important thing is to obtain meaningful information from the test without dealing with complex statistical procedures and Rasch measurement has this capability. Certain areas such as early childhood and special education are struggling to find a way to assess child development or to identify learning disability qualification. There is a large gap between people in educational measurement/statistics and practical filed that need to be bridged. A Rasch-assisted report form or the map of the test is an easy way to achieve that goal.

Futoshi Yumoto

IOMW XII
Cairns, Australia
June 28 - July 3, 2004

Ready for a little temptation? Has IOMW ever offered such lush location? Try this:

www.soe.jcu.edu.au/IOMW2004/

the website for *IOMW Down Under!* None of the photos on the IOMW site are promo glossies; just snapshots I took one weekend. IOMW will have fun and substance! Many key folk in the Rasch measurement have undertaken to participate. The program already includes paper sessions, software workshops, networking opportunities, and SCUBA diving. This is the one you'll tell your grand-kids about!

Joseph Indorato, Operations Manager, at Quest Marlin Cove *www.quest-inns.com.au,*is waiting to help you with your accommodation requirements and is ready to take those early bookings:

jindorato@questapartments.com.au

Joe is willing to help you share an apartment if you give him the appropriate information. Remember to mention the magic code *IOMW* when booking/enquiring. Quest Inns also has accommodation all round Australia for those who want to make the most of the visit.

Trevor Bond, Chair IOMW XII Committee

Journal of Applied Measurement
Volume 4, Number 1. Spring 2003

The Effect of Missing Data on Estimating a Respondent's Location using Ratings Data. *R. J. De Ayala*

Rasch Simultaneous Vertical Equating for Measuring Reading Growth. *Lee Ong Kim*

An Examination of Exposure Control and Content Balancing Restrictions on Item Selection in CATs using the Partial Credit Model. *Laurie Laughlin Davis, Dena A. Pastor, Barbara G. Dodd, Claire Chiang and Steven J. Fitzpatrick*

Aerobic Exercise Equipment Preferences among Older Adults: A Preliminary Investigation. *Marilyn A. Looney and James H. Rimmer*

Measurement Precision of the Clinician Administered PTSD Scale (CAPS): A Rasch Model Analysis. *Elizabeth J. Betemps, Richard M. Smith, Dewleen G. Baker, and Barnara A. Rounds-Kugler*

A Comparative Evaluation of Methods of Adjusting GPA for Differences in Grade Assignment Practices. *Pui-Wa Lei, Dina Bassiri, and E. Matthew Schulz*

Understanding Rasch Measurement: An Introduction to Multidimensional Measurement. *Derek C. Briggs and Mark Wilson*

This issue marks the transition to an expanded format for the journal. It has a new page size and a new look (double column format to make it easier to read), as well as expanding to seven articles per issue. Sample copies will be available at the JAM booth in the exhibit hall at the AERA annual meeting in Chicago.

For subscriptions, submissions, back-issues and instructors' sample copies, contact:
Richard M. Smith, Editor
Journal of Applied Measurement
P.O. Box 1283, Maple Grove, MN 55311
JAM web site: http://home.att.net/~rsmith.arm

"... for only by varied iteration can alien conceptions be forced on reluctant minds."
Herbert Spencer, in the preface to *The Data of Ethics,* 1881. *Courtesy of Lise DeShea*

Rasch Measurement Transactions
P.O. Box 811322, Chicago IL 60681-1322
Tel. & FAX (312) 264-2352
rmt@rasch.org www.rasch.org/rmt/
Editor: John Michael Linacre

SIG Chair: Trevor Bond SIG Secretary: Ed Wolfe

Rasch-related Sessions at AERA Annual Meeting, Chicago

Monday, April 21, 2003

D2-12 Methods in Educational Measurement and Statistics: Poster Fair
12:00 p.m. - 1:30 p.m., Hyatt, Riverside Center, Exhibition Hall

Three Exposure Control Techniques in CAT Using the Generalized Partial Credit Model
Winona Burt, The University of Texas at Austin/Evaluation Software Publishing, Inc.; Soo-jin Kim, The University of Texas at Austin; Laurie Davis, NCS Pearson; Barbara Dodd, The University of Texas at Austin

Comparison of NOHARM and Conditional Covariance Methods of Dimensionality Assessment
Holmes Finch, Brian Habing, Huynh Huynh, University of South Carolina

The Evaluation of Exposure Control Procedures for an Operational CAT
Brian French, Purdue University; Tony Thompson, ACT Inc.

Ability Distribution Recovery Under the 3PL Model With Unknown Item Parameters
Yaowen Hsu, Meichu Fan, ACT Inc.

TIMSS 1999 Factor Invariance Across U.S. Samples of Males and Females
Jason Immekus, Susan Maller, Purdue University

The comparison of Angoff probability ratings with actual item performance: why does Angoff not work?
Rianne Janssen, Paul De Boeck, University of Leuven

A Crossed Multilevel IRT Model for Analyzing Rated Polytomous Measures
Sang-Jin Kang, Yonsei University; Mark Wilson, University of California, Berkeley

Applications of the Continuation Ratio Model for Ordered Category Items
Seock-Ho Kim, The University of Georgia

Evaluating a New Approach to Detect Aberrant Responses in CAT
Ying Lu, University of Massachusetts Amherst; Frederic Robin, Educational Testing Service

Reconsidering reliability in a multi-level context
Stuart Luppescu, University of Chicago; Robert Gladden, Consortium on Chicago School Research; Anthony Bryk, University of Chicago

Ability estimation under different item parameterization and scoring models
Randall Schumacker, University of North Texas; Benjamin Si, Hong Kong; Mount Robert, Dallas ISD

D2-13 Methods in Educational Research: Poster Fair
12:00 p.m. - 1:30 p.m., Hyatt, Riverside Center, Exhibition Hall

The relationship between Teacher Assessment and Pupil Attainments in Standard Test Tasks at Key Stage 1, 1997-2000.
Iasonas Lamprianou, Bill Boyle, CFAS, Faculty of Education, Univ. of Manchester, UK

Characteristics of Mathematics Items Associated with Gender DIF
Yanmei Li, Allan Cohen, University of Wisconsin-Madison; Robert Ibarra, University of New Mexico

April 2003, Chicago

April 19-20, Saturday-Sunday
An Introduction To Rasch Measurement: Theory And Applications.
Workshop at the University of Illinois at Chicago conducted by Everett V. Smith Jr. and Richard M. Smith. 312/996-5630
evsmith@uic.edu

April 21-25, Monday-Friday
AERA Annual Meeting. *www.aera.net*

April 25-27, Friday-Sunday
Ben Wright *Festschrift.* *www.rasch.org*

April 28-29, Monday-Tuesday
***Facets* Workshop,** CORE, Evanston
www.winsteps.com/seminar.htm

April 30-May 1, Wednesday-Thursday
***Winsteps* Workshop**, CORE, Evanston
www.winsteps.com/seminar.htm

Rasch analysis of inattentive, hyperactive and impulsive behavior in young children and the link with academic achievement
Christine Merrell, Peter Tymms, University of Durham

Can Judges Identify Easy and Difficult Questions Holistically in a Standard Setting Study?
Susan Thomas, IBM Corp

Collecting Ethnic and Racial Data: Holding Steady as Demographics Change
Cathy Wendler, Educational Testing Service; Jose-Felipe Martinez-Fernandez, University of California Los Angeles; Robin Hochman, Anna Kubiak, Educational Testing Service

D1-17 Discussion on Measurement: Paper Discussion
2:15 p.m. - 2:55 p.m., Hyatt, Grand Ballroom E, East Tower - Gold Level

An Application of a Special Two-Class Item Response Model Using Markov Chain Monte Carlo Method
Yiyu Xie, University of California at Berkeley

Parameter Estimation Under HGLM Versus IRT for Polytomous Items
Natasha Williams, S. Natasha Beretvas, University of Texas at Austin

Evaluating Computer-based Testing Security by Generalized Item Overlap Rates
Jinming Zhang, Ting Lu, ETS

Scaling early reading performance in state testing programs
Paul Yovanoff, Gerald Tindal, University of Oregon

SIG-LOS-1 Analyzing Change in Educational Settings: Paper Session
2:15 p.m. - 3:45 p.m., Hyatt, Grand Suites 1, East Tower - Gold Level

A repeated measures, multilevel Rasch model with application to self-reported criminal behavior
Christopher Johnson, Steve Raudenbush, University of Michigan

Tuesday, April 22, 2003

D1-25 Score Stability, Distribution, Efficacy and Augmentation: Paper Session
8:15 a.m. - 10:15 a.m., Hyatt, Regency A, West Tower - Gold Level

A Comparison of Item- and Testlet-Level Scoring on Scale Stability in the Presence of Test Speededness
James Wollack, Craig Wells, Allan Cohen, UW-Madison

Investigating Constructed Response Scoring Over Time: The Effects of Study Design on Trend Rescore Statistics
Melinda Hess, University of South Florida; John Donoghue, Educational Testing Service

Comparing the efficacy of generalizability theory-based and IRT-based domain scores in a matrix sampling environment
Deborah Harris, ACT Inc.; Bradley Hanson, CTB; Xiaohong Gao, ACT Inc.

D1-28 Psychometric Potpourri: Paper Session
8:15 a.m. - 10:15 a.m., Hyatt, Columbus Hall G/H, East Tower - Gold Level

A comparative investigation of analyzing sources of variation in the observational rating system
Sungsook Kim, UC Berkeley

A fit test for the Rasch model based on Monte-Carlo simulation
Judit Antal, The Ohio State University

D1-18 New Member Session on Measurement: Poster Session
12:25 p.m. - 1:55 p.m., Hyatt, Riverside Center, Exhibition Hall

The Use of Person-Fit Statistics to Evaluate Placement Tests
Hamzeh Dodeen, United Arab Emirates University

The Construct Validity of Scores on Self-Efficacy and Self-Concept Measures
Nai-Kuang Ku, University of Southern California

Differential prediction bias in the Wide Range Intelligence Test across race, gender, and education level
Jennifer Shields, Timothy Konold, University of Virginia; Joseph Glutting, University of Delaware

Minicourse F: . Interpreting Rasch Model Fit Statistics.
1 p.m. - 5 p.m., Fairmont Hotel, Gold (admission by ticket only, $30. through AERA registration).
See *www.aera.net/meeting/am2003/courses/ProfDevCourses03.pdf*
John M. Linacre, University of the Sunshine Coast; Thomas R. O'Neill, University of Illinois at Chicago

Wednesday, April 23, 2003

D1-26 Multidimensional Models, Equating and Reliability: Paper Session
8:15 a.m. - 10:15 a.m., Hyatt, Regency B, West Tower - Gold Level

Comparison of Multidimensional IRT Equating Methods with Small Samples
Kyung-Seok Min, Michigan State University; Jong-Pil Kim, ACT Inc.

Sensitivity of IRT equating on the behavior of test equating items
Michalis Michaelides, Stanford University

Evaluating the Accuracy of Pre-equated Test Forms
Thomas Langenfeld, Chi-Yu Huang, Judith Spray, ACT Inc; Charles Kunce, National Institute for Automotive Service Excellence

Modeling Mathematics Problem Solving Item Responses Using a Multi-dimensional IRT model
Margaret Wu, Ray Adams, University of Melbourne

Estimating Scale Reliability of Multidimensional Composite Scores
Akihito Kamata, Ahmet Turhan, Eqbal Darandari, Florida State University

D1-31 Research on Calibration and Estimation: Paper Session
Wednesday, 8:15 a.m. - 10:15 a.m., Hyatt, Regency C, West Tower - Gold Level

The Effects of Multidimensional Polytomous Response Data on Unidimensional Many-FACET Rasch Model Parameter Estimates
Shudong Wang, Ning Wang, CAT*ASI

Expected values and reliability of number-right scores for IRT calibrated items
Dimiter Dimitrov, Kent State University, Kent, Ohio

Online Calibration And Scale Stability Of A CAT Program
Fanmin Guo, Lin Wang, ETS

Thursday, April 24, 2003

SIG-RM-4 Rasch Model Scaling: Paper Session
8:15 a.m. - 10:15 a.m., Sheraton, Columbus B, Level 3

Re-examining the Quantitative Imperative
William Fisher, LSU Health Sciences Center

Analyzing DIF in Polytomous Responses of University Alumni to a Follow-up Questionnaire
Joseph Curtin, Richard Sudweeks, Brigham Young University; Richard Smith, American Institutes for Research

A Rasch Measurement Example in Grant Application Process
Yesim Capa, William Loadman, The Ohio State University

A Rasch-derived scale for measurement of strength of motivation for medical training
Ronny Wierstra, Dept of Educational Sci Utrecht; Marja Nieuwhof, Olle Ten Cate, University Medical Center Utrecht

Examination of Rasch structure of feminine identity scales
Larry Ludlow, James Mahalik, Boston College; Camelia Rosca

Scaling Resident Assistant Effectiveness
Christine Mills, Boston College

Satisfaction With Organization/Management
Johnna Gueorguieva, Donna Tatum, American Society for Clinical Pathology

Chair: Michael Yoes, NCS Pearson. Discussant: Christa Winter, Springfield College

D1-24 IRT Item Parameter Estimation: Paper Session
10:35 a.m. - 12:05 p.m., Sheraton, Huron, Level 2

Missing Data and IRT Item Parameter Estimation
Christine DeMars, James Madison University

An Empirical Investigation of the Hybrid IRT Model for Improving Item Parameter Estimation in Speeded Tests
Daniel Bolt, Andrew Mroch, Jee-Seon Kim, University of Wisconsin, Madison

A Long-Term Study of the Stability of Item Parameter Estimates
G. Gage Kingsbury, NWEA

Impact of Item Drift With Non-normal Distributions
Elizabeth Witt, John Stahl, Betty Bergstrom, Robin Ingalls, Promissor

SIG-RM-1 Recent Progress in Rasch Measurement Theory: Paper Session
4:05 p.m. - 6:05 p.m., Hyatt, Horner, Silver Level

What are the Effects of allowing Crossing Item Characteristic Curves into our Measurement Model?
Timothy Pelton, University of Victoria

Modeling Local Item Dependence Using the One-Parameter Hierarchical Generalized Linear Model
Hong Jiao, Akihito Kamata, Florida State University

Decomposition of Rasch Partial Credit Items and Applications
Huynh Huynh, J. Patrick Meyer, University of South Carolina

Reliability of true cutting scores for Rasch calibrated items
Dimiter Dimitrov, Kent State University, Kent, Ohio

A confirmatory approach to dimensionality assessment.
Michel Fournier, Jean-Guy Blais, University of Montreal

Structure and Invariance: Using Wright's Method and Confirmatory Factor Analysis in Assessing Change Over Time
Kathy Green, Thomas Paskus, University of Denver; Cynthia Jew, California Lutheran University; Diana Stephens, University of Redlands

Chair: Randall Schumacker, University of North Texas. Discussant: E Schulz, ACT Inc.

SIG-RM-2 Rasch Measurement SIG: Business Meeting
6:15 p.m. - 7:45 p.m., Hyatt, Horner, Silver Level

A Trade-off Between Consistency of Responses and Precision of Measurement
David Andrich, Murdoch University

News of the Rasch Measurement SIG
George Karabatsos, University of Illinois-Chicago

Friday, April 25, 2003

SIG-RM-3 Rasch Measurement: Issues And Practice: Paper Session
8:15 a.m. - 10:15 a.m., Hyatt, Wright, Silver Level

Reliability in Rasch Measurement: Avoiding the Rubber Ruler
Randall Schumacker, University of North Texas

An Evaluation of the Multi-faceted Rasch Model in Analyzing Job Task Survey Data
Ning Wang, Shudong Wang, CAT*ASI

Testing Equating with the presence of DIF
Kwanglee Chu, Akihito Kamata, Florida State Univ.

The effect of markers and optional questions on the results of high-stakes exams
Iasonas Lamprianou, CFAS, Faculty of Education, Univ. of Manchester, UK; Kyriakos Pillas, Research and Evaluation Unit, Pedagogical Institute of Cyprus

Using Displacement Criteria to Anchor Multiple Choice Tests
Surintorn Suanthong, Mary Lunz, Measurement Research Associates, Inc.

Failing Standards: Validity Concepts in Criterion-Referencing
Gregory Stone, University of Toledo

Chair: Gene Kramer, American Dental Association. Discussant: Dorthea Juul, American Board of Psychiatry and Neurology

SIG-ES-11 Issues in Multivariate Statistics and Psychometrics: Paper Session
8:15 a.m. - 10:15 a.m., Sheraton, Michigan B, Level 2

Modeling Longitudinal Ordinal Response Variables for Educational Data
Ann O'Connell, D. Betsy McCoach, University of Connecticut

A Corrected Asymptotic Distribution of an IRT Fit Measure that Accounts for the Effects of Item Parameter Estimation
John Donoghue, Catherine Hombo, Educational Testing Service

FESTSCHRIFT in honor of Ben Wright

25-27 April 2003 (weekend after AERA)
Rehabilitation Institute of Chicago
345 E. Superior Street, Chicago. Heyworth Rooms - 2nd floor (south end)

"Access, Provocation, and the Development of Professional Identity: Celebrating the Careers of Benjamin D. Wright."

Celebrate Ben Wright with friends and colleagues through
1) platform presentations;
2) poster presentations, roundtables, and "artifact" displays;
3) software demonstrations (Friday afternoon, 25 April), and
4) a social event (Saturday evening, 26 April).
The conference will close by early afternoon on Sunday.

The program is coming together nicely with commitments from many of Ben's closest colleagues and students, but **it is not too late to send in presentation proposals.** Those intending to be there should go to **www.rasch.org** to obtain, print, and fax/mail/email in a **free registration form** so we have some idea of how many will be showing up.

An Introduction To Rasch Measurement: Theory And Applications

April 19-20, 2003 (Saturday and Sunday)

This **Training Session** will be presented at the University of Illinois at Chicago immediately before the AERA Annual Meeting in Chicago. The workshop will be conducted by Everett V. Smith Jr., PhD (University of Illinois at Chicago) and Richard M. Smith, PhD (Educational Data Systems). Registration information is available from Everett Smith at 312/996-5630 or *evsmith@uic.edu* - **Fee:** $250.00 ($100 for students with proof of student status).

Registration deadline: April 1, 2003

Workshop Description: The purpose of this training session is to introduce participants to the theory and applications of Rasch measurement and provide hands-on experience using Rasch calibration programs to scale ordinal data. This session 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 self-contained units. The units are: 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 Analysis; and Analysis of Participants Data. The co-directors will divide the topics in each session to maximize individual strengths. The instructional format will combine lecture, question and answer, and small group instruction.

Saturday, April 19, 2003

8:30 am **Registration and Coffee/Juice/Danishes/Muffins**

Session I Introduction to Rasch Measurement

9:00 **Welcome:** Richard M. Smith
9:05 **What is Measurement:** Richard M. Smith
9:30 **Rasch Measurement Models:** Everett V. Smith, Jr.
10:00 **True Score vs. Rasch Measurement Models:** Everett V. Smith, Jr.
10:15 **Break**

Session II Item and Person Calibration

10:45 **Testing the Fit of Data:** Richard M. Smith
11:30 **Dimensionality and PC Analysis of Residuals**: Everett V. Smith, Jr.
12:00 pm **Lunch Break**

Session III Dichotomous and Polytomous Data

1:30 **WINSTEPS Control Language:** Everett V. Smith, Jr.
2:00 **Small Group Calibration Projects:**
Group 1 Dichotomous Data. Richard M. Smith
Group 2 Polytomous Data. Everett V. Smith, Jr.
2:45 **Break**

Session IV Performance and Judged Data

3:15 **FACETS Control Language:** Richard M. Smith
3:45 **Small Group Calibration Projects**
Group 1 Nested Data (ratings of conference proposals). Richard M. Smith
Group 2 Fully Crossed Data (ratings of student performance). Everett V. Smith, Jr.

4:30 **End of Day One**

5:30 Optional group dinner

Sunday, April 20, 2003

8:30 am **Coffee/Juice/Danishes/Muffins**

Session V Applications of Rasch Measurement

9:00 **Score Reporting:** Everett V. Smith, Jr.
9:25 **Standard Setting:** Everett V. Smith, Jr.
9:50 **Item Bias** : Richard M. Smith
10:15 **Break**

Session VI Applications of Rasch Measurement

10:45 **Test Equation and Item Banking:** Richard M. Smith
11:15 **Computer Adaptive Testing:** Richard M. Smith
11:45 **Rasch vs. Multi-Parameter IRT Models**: Everett V. Smith, Jr.
12:00 pm **Lunch Break**

Session VII Examples of Rasch Analysis

1:30 **Rating Scale Data:** Everett V. Smith, Jr.
2:00 **Partial Credit Data:** Richard M. Smith
2:30 **Break**

Session VIII Analysis of Participants Data

3:00 **Running WINSTEPS and FACETS**: Everett V. Smith, Jr., Richard M. Smith

3:30 **Your turn to analyze data: Participants interested in analyzing their own data should bring a laptop** with Windows 95 (or newer), 8 MB RAM (min), and Wordpad or Notepad.

4:30 **End of Workshop**

Estimating 50% Cumulative Probability Thresholds

Rating scale category boundaries can be conceptualized in a number of ways. L. L. Thurstone (1928) describes the computation of .50 cumulative proportions as "scale values." These scale values are now referred to as Thurstone thresholds. They are also the parameters in the "Graded Response" model.

Rasch rating scale structures are parameterized using the points of equal-probability of adjacent categories, rather than the points of equal probability of accumulated category probabilities. Nevertheless, in communicating Rasch findings, it can be convenient to represent Rasch rating scale functioning in terms of Thurstone-type thresholds.

Rasch polytomous models, such as Andrich "rating scale" or Masters "partial credit" models have the form:

$$\log\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) \equiv B_n - D_i - F_{gj}$$

with the usual notation conventions, and $F_{g0} \equiv 0$. F_{gj} parameterizes the "Rasch threshold" or "step", the point of equal probability of categories j-1 and j. The subscript "g" indicates the manner in which the set of $\{F_{gj}\}$ parameters relates to the *n* or *i* parameters. For the Andrich "Rating Scale" model, "g" signifies all items. For the Masters' "Partial Credit" model, "g" signifies item *i.* For Ben Wright's "Style" model, "g" signifies person *n.* For an instrument in which different groups of items share common rating scales, "g" identifies the item groups.

Let the Thurstone-type thresholds be identified as $\{T_{gj}\}$ relative to item difficulty, D_i. Then

$$\sum_{k=0}^{j-1} P_{(Tgj)k} = 0.5 = \sum_{k=j}^{m} P_{(Tgj)k}$$

for j=1,m. So that, multiplying through by the normalizer,

$$\sum_{k=0}^{j-1} e^{kT_{gj} - \sum_{h=0}^{k} F_{gh}} = \sum_{k=j}^{m} e^{kT_{gj} - \sum_{h=0}^{k} F_{gh}}$$

Then let $t_j = e^{T_{gj}}$ and $c_k = e^{-\sum_{h=0}^{k} F_{gh}}$

so that, for each of j=1,m,

$$1 + \sum_{k=1}^{j-1} t_j^{\,k} . c_k - \sum_{k=j}^{m} t_j^{\,k} . c_k = 0$$

If the T_{gj} are specified, then the t_j are known, and the c_k can be obtained by solving the *m* simultaneous equations. From the c_k, the F_{gh} can be computed directly. Thus a polytomous Rasch model can be parameterized in terms of Thurstone-type thresholds using matrix notation and Cramer's rule. On the other hand, if the F_{gh} are specified, then the c_k are known. Each of the *m* equations becomes a polynomial in t_j. The required root always exists. The lower bound of the search for t_j is zero (when the polynomial must be positive), and t_j can be increased until the polynomial becomes negative. When the value of t_j has been found for which the equation is well enough satisfied, then T_{gj} is computed.

A 3 category, so two threshold, item has Rasch thresholds –0.85, 0.85. The lower Thurstone-type Threshold is given by:

$$1 - t_1 e^{0.85} - t_1^{\,2} = 0$$

so that $t_1 = 2.7$, and T_1 is -1.0. By symmetry, T_2 is +1.0.

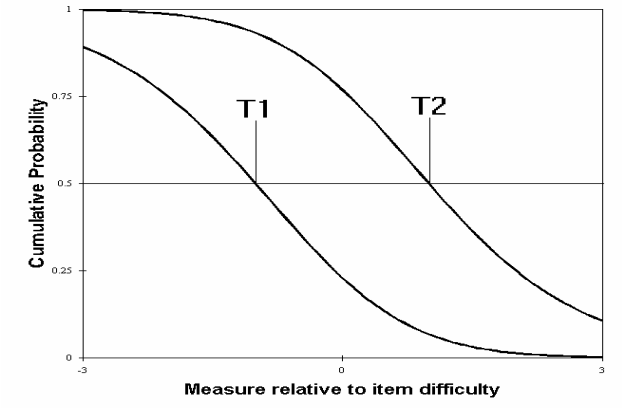
Working backwards for the Rasch thresholds, if the Thurstone-type thresholds are -1, +1, then

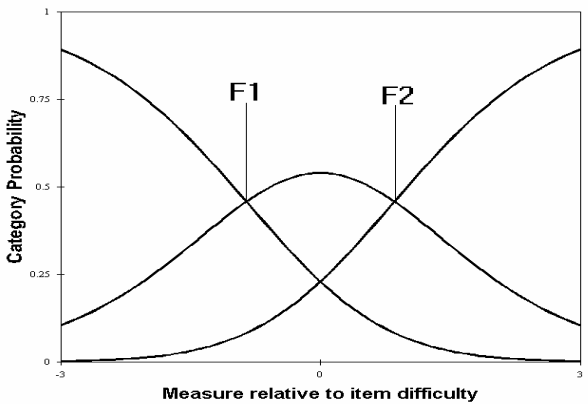
$$1 - e^{-1} e^{F_1} - e^{-2} e^{0} = 0$$

So that $F_1 = \log(1/(e^1 - e^{-1})) = -.85$, and $F_2 = +.85$.

John M. Linacre

Thurstone L.L. (1928) Attitudes can be measured. *American Journal of Sociology, 33,* 529-54.





What Is Good About Rasch Measurement?

Reprinted with permission from *www.personalityresearch.org/psychometrics/measurement.html*

What is good about Rasch measurement (Rasch, 1960/1980)? (The Rasch model is a one-parameter logistic model within item response theory in which a person's level on a latent trait and the level of various items on the same latent trait can be estimated independently yet still compared explicitly to one another.) To answer this question requires a specification of what is meant by *measurement*. Two main approaches to defining measurement are the traditional approach and the representational approach. The traditional approach has been widely accepted in the physical sciences since its development by Holder (1901, as cited in Michell, 1997), who synthesized the approaches of Euclid, Newton, and Dedekind. One feature of the traditional approach is that it entertains an empiricist account of number. Measurement, then, becomes the exercise of establishing a correspondence between quantitative variables in the world and numerical instruments (Mill, 1843/1973). Mill's empiricist conception of number was criticized and made to seem untenable by Frege (1884/1984), and the traditional approach to measurement was criticized by Russell (1903), who developed a representational theory of measurement. The representational theory was further advanced by Krantz, Luce, Suppes, and Tversky (1971), Suppes, Krantz, Luce, and Tversky (1989), and Luce, Krantz, Suppes, and Tversky (1990), who are its most sophisticated contemporary proponents. Michell (1994, 1997), on the other hand, has emerged as the most sophisticated contemporary proponent of the traditional theory of measurement.

On both the traditional theory and the representational theory, Rasch measurement is good because it is an example of additive conjoint measurement. Rasch measurement satisfies two conditions that are necessary in order for an attribute to be quantitative. First, the attribute must possess additivity. Second, the attribute must possess ordinality. The Rasch model possesses additivity because the difference between the manifest level and the latent level involves the additive measurement of two different latent variables – one for the person, one for the item. The Rasch model possesses ordinality because person and item variables can be explicitly compared at the latent level as being higher or lower than one another.

Rasch measurement is good partly because it stands in contrast to a ridiculous version of the representational theory that has gained ascendance within psychology: namely, operationism – that is, the idea that a variable is completely defined by the operations or measurements used to recognize it. There may be some ontological differences between the traditional theory and the Krantz et al. representational theory regarding the state of the world, whether the variables to be measured are quantitative or qualitative, but neither of these theories is completely subjective and idealistic (in the Berkeleyan sense) in the way that operationism is. Operationism permits quantification of anything whatsoever, albeit in a wholly arbitrary way. Operationism thus exemplifies a strong Pythagorean tendency within psychology, supposing as it does that numbers can be applied to anything. The operations used to generate the numbers, however, may represent nothing other than themselves. Operationism justifies applying a rule – any rule – to empirical reality. Applying a different rule may result in a different result, but both rules are right by fiat, because they define what they purport to represent.

It may be a mistake to claim that Rasch measurement is an idealization. According to the traditional theory, an idealization is not measurement. Within any given application, however, measurement may be impossible. Indeed, Kant (1786/1970) and Searle (1994) seem to think that psychological variables such as consciousness are inherently non-quantitative. For the representational theory, applying numbers to a qualitative reality in a systematic and rigorous way is the model for measurement. Thus, the representational theory entails no quivering reflections on whether psychology can ever be a quantitative science. The traditional theory does entail such reflections, however, because, within this theory, whether any given attribute is quantitative is an empirical question to which the answer may be "no."

If psychological variables turn out to be non-quantitative, this does not entail that psychology cannot be a science. First, psychological variables will continue to have predictive power and thus practical utility. The correlation between any number of psychological traits and criterion variables, for example, ranges from .3 to .5 (Mischel [1968] has called this the *personality coefficient*). Explanations for these regularities, however, will have to be acknowledged as being speculative and theoretical, bringing psychology into close alliance with philosophy. The search for quantitative variables, however, may represent the wave of the future for a potentially quantitative scientific psychology.

G. Scott Acton
University of California, San Francisco.

References:

Frege, G. (1984). *The foundations of arithmetic* (J. L. Austin, Trans.). New York: Blackwell and Mott. (Original work published 1884)

Kant, I. (1970). *Metaphysical foundations of natural science* (J. Ellington, Trans.). Indianapolis, IN: Bobbs-Merrill. (Original work published 1786)

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Vol. 1. Additive and polynomial representations.* New York: Academic.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement: Vol. 3. Representation, axiomatization, and invariance.* San Diego, CA: Academic.

Michell, J. (1994). Numbers as quantitative relations and the traditional theory of measurement. *British Journal for the Philosophy of Science, 45,* 389-406.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88,* 355-383.

Mill, J. S. (1973). *A system of logic*. Toronto, Canada: University of Toronto Press. (Reprinted from *Collected works of John Stuart Mill, Vol. 7*, by J. M. Robson, Ed.). (Original work published 1843)

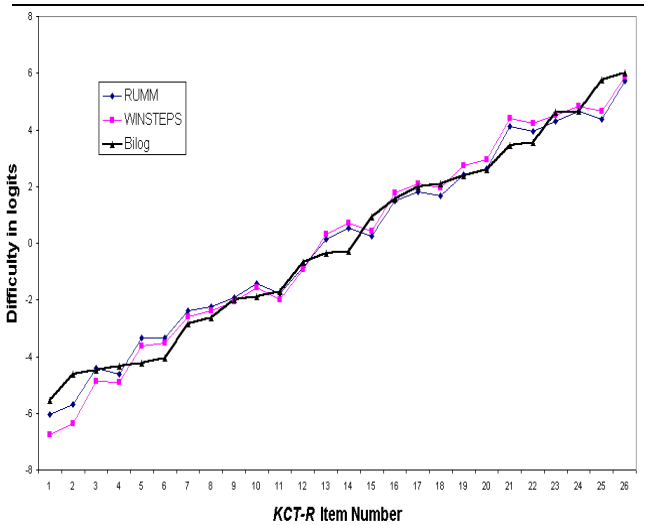
Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests (expanded ed.).* Chicago: The University of Chicago Press. (Original work published 1960).

Russell, B. (1903). *Principles of mathematics.* New York: Cambridge University Press.

Searle, J. R. (1994). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

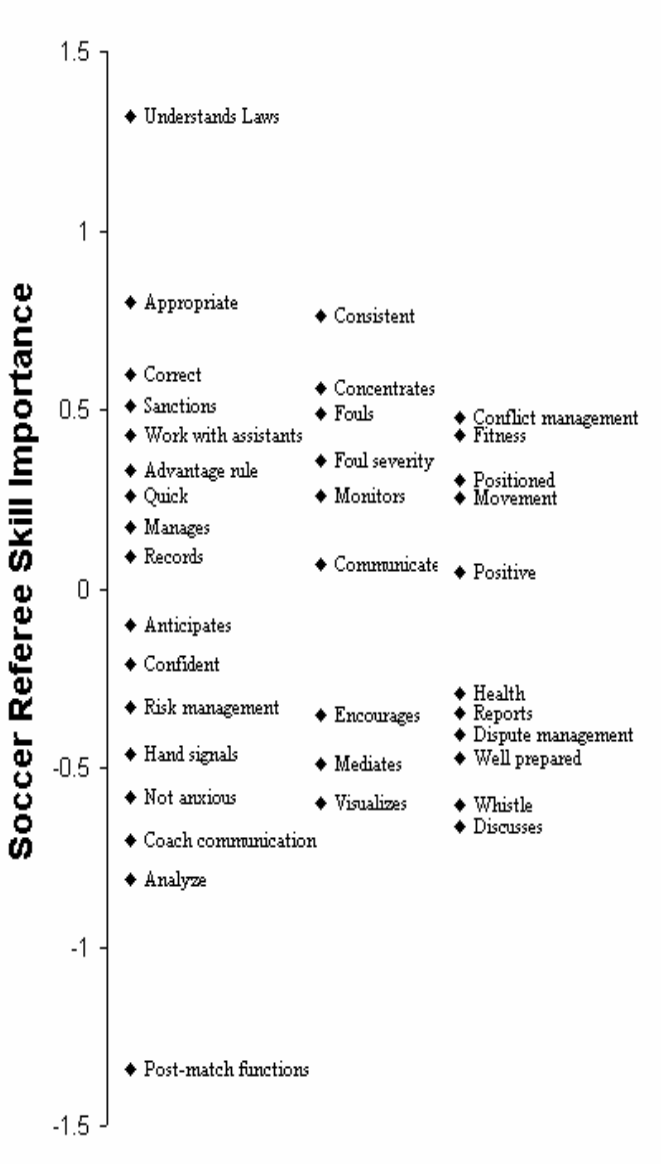
Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement: Vol. 2. Geometrical, threshold, and probabilistic representations.* San Diego, CA: Academic.



Comparison of Rasch item estimates by Mark Stone & Futoshi Yumoto, 2002. *Presented at COMET, Chicago, Feb. 2003. An article and description are forthcoming.*

2nd International Conference
on Measurement in Health, Education, Psychology and Marketing: Developments with Rasch models
Perth and Fremantle, Western Australia
January 20 –22, 2004

Exciting developments in the theory and practice of measurement in health, education, psychology and marketing provide an opportunity to review the state of the art in measurement science, learn from the experts in an extensive pre-conference program, and enjoy the delights of summer in Western Australia. **Abstracts by July 31, 2003.** For further information, email Angelina Chillino, ***chillino@murdoch.edu.au***



Soccer Referee Skill Importance map constructed using measures in "CriLT Project Publications: Elite Refereeing – Soccer Specific" (n.d.) by Scott Dickson, University of New England, Armidale, Australia.

The Zero-Discrimination Paradox

The dichotomous Rasch model specifies that all items have the same discrimination. But what happens if that discrimination is zero? Some critics perceive here a flaw in the Rasch model, but, paradoxically, the Rasch model analysis is accurate. It is the 2-PL analysis that is flawed!

Here is a 2-PL IRT model including its usual item discrimination parameter, a_i:

$$\log\left(\frac{P_{ni}}{1-P_{ni}}\right) \equiv a_i(\theta_n - b_i)$$

If all item discriminations are the same, a_i=a and this becomes a Rasch model. Thus data which fit a 2-PL model with uniform item discriminations also fit a Rasch model with Rasch parameters $B_n = a.\theta_n$ and $D_i = a.b_i$. This presents no conceptual difficulties except in the case of a=0. A 2-PL analysis would, one imagines, report that a=0, but the Rasch analysis cannot do this, so what would it report?

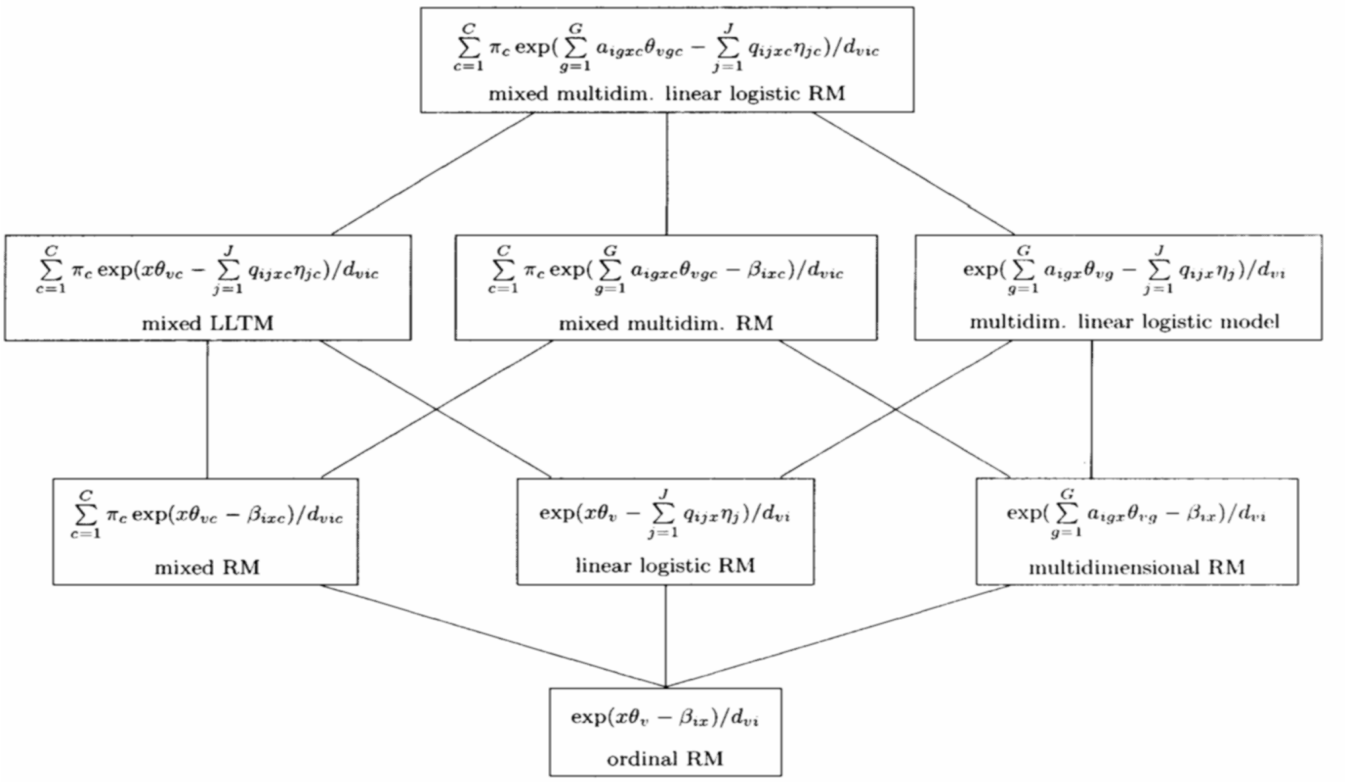
If all a_i=0 in the 2-PL model statement above, then P_{ni} = 0.5 for all *n* and *i.* So that, from the Rasch perspective, all B_n = B and all D_i = D, and B=D. This is equivalent to coin-tossing. The Rasch analysis would unambiguously report that all person abilities equal all item difficulties, and the data would fit the Rasch model.

In fact, a Rasch analysis can go further. If the items are not discriminating, so that B_n=B and D_i=D, but **B≠D**, then this is equivalent to tossing a **biased** coin. B-D is a measure of the bias in the coin. Rasch would report correctly that all person abilities are equal, and that all item difficulties are equal, but that person ability is unequal to item difficulty. The data would fit the model.

With a biased coin, 2-PL estimation algorithms encounter a paradox. If item discrimination dominates, then a_i=a=0 is reported, but the resultant model does not fit the data. This is because a=0 implies P_{ni}=0.5, but in fact P_{ni}≠0.5.

If ability and difficulty dominate, then $\theta_n=\theta$ and b_i=b and a_i=a=1 (or a constant, not equal to 0). The model does fit the data, but 2-PL now misreports the uniform zero discrimination as non-zero! In a situation in which the Rasch measures are straightforward to interpret, it is the 2-PL estimates that are either incorrect or misleading.

John M. Linacre



Jürgen Rost's "hierarchy of generalized Rasch models" from "The growing family of Rasch models" (Chapter 2 in "Essays on Item Response Theory", A. Boomsma et al. (Eds.), New York: Springer. 2000). "Ordinal RM" includes dichotomous, rating scale, and partial credit models. "Linear logistic RM" includes many-facet models. The SALTUS model is a "mixed LLTM" model. "Multidimensional RM" would include Georg Rasch's own "generalization to the case of more responses than one" in "An Individual-Centered Approach to Item Analysis with Two Categories of Answers" (1964).