Rasch with an Error Term

$$X_{ni} = P_{ni} \pm \sqrt{(P_{ni}*(1-P_{ni}))}$$

$$X_{ni} = E_{ni} \pm \sqrt{(W_{ni})}$$

RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 23 No. 4 Spring 2010 ISSN 1051-0796

Statistics and Measurement: Clarifying the Differences

Measurement is qualitatively and paradigmatically quite different from statistics, even though statistics obviously play important roles in measurement, and *vice versa.* The perception of measurement as conceptually difficult stems in part from its rearrangement of most of the concepts that we take for granted in the statistical paradigm as landmarks of quantitative thinking. When we recognize and accept the qualitative differences between statistics and measurement, they both become easier to understand.

Statistical analyses are commonly referred to as quantitative, even though the numbers analyzed most usually have not been derived from the mapping of an invariant substantive unit onto a number line. Measurement takes such mapping as its primary concern, focusing on the quantitative meaningfulness of numbers (Falmagne & Narens, 1983). Statistical models focus on group processes and relations among variables, while measurement models focus on individual processes and relations within variables (Duncan, 1992). Statistics makes assumptions about factors beyond its control, while measurement sets requirements for objective inference (Andrich, 1989). Statistics primarily involves data analysis, while measurement primarily calibrates instruments in common metrics for interpretation at the point of use (Cohen, 1994).

The scientific value of statistics resides largely in the reproducibility of cross-variable data relations. Statistics focuses on making the most of the data in hand, while measurement focuses on using the data in hand to inform (a) instrument calibration and improvement, and (b) the prediction and efficient gathering of meaningful new data on individuals in practical applications. Where statistical "measures" are defined inherently by a particular analytic method, measures read from calibrated instruments - and the raw observations informing these measures - need not be computerized for further analysis.

Because statistical "measures" are usually derived from ordinal raw scores, changes to the instrument change their meaning, resulting in a strong inclination to avoid improving the instrument. Measures, in contrast, take missing data into account, so their meaning remains invariant over instrument configurations, resulting in a firm basis for the emergence of a measurement quality improvement culture. So statistical "measurement" begins and ends with data analysis, where measurement from calibrated instruments is in a constant cycle of application, new item calibrations, and critical recalibrations that require only intermittent resampling.

The vast majority of statistical methods and models make strong assumptions about the nature of the unit of measurement, but provide either very limited ways of checking those assumptions, or no checks at all. Statistical models are descriptive in nature, meaning that models are fit to data, that the validity of the data is beyond the immediate scope of interest, and that the model accounting for the most variation is regarded as best. Finally, and perhaps most importantly, statistical models are inherently oriented toward the relations among variables at the level of samples and populations.

Measurement models, however, impose strong requirements on data quality in order to achieve the unit of measurement that is easiest to think with, one that stays constant and remains invariant across the local particulars of instrument and sample. Measurement methods and models, then, provide extensive and varied ways of checking the quality of the unit, and so must be prescriptive rather than descriptive. That is, measurement models define the data quality that must be obtained for objective inference. In the measurement paradigm, data are fit to models, data quality is of paramount interest, and data quality evaluation must be informed as much by qualitative criteria as by quantitative.

To repeat the most fundamental point, measurement models are oriented toward individual-level response processes, not group-level aggregate processes. Herbert Blumer pointed out as early as 1930 that quantitative

Table of Contents

method is not equivalent to statistical method, and that the natural sciences had conspicuous degrees of success long before the emergence of statistical techniques (Hammersley, 1989). Both the initial scientific revolution in the 16th-17th centuries and the second scientific revolution of the 19th century found a basis in measurement for publicly objective and reproducible results, but statistics played little or no role in the major discoveries of the times. Now we are in a position to appreciate a comment by Ernst Rutherford, the winner of the 1908 Nobel Prize in Chemistry, who held that, if you need statistics to understand the results of your experiment, then you should have designed a better experiment (Wise, 1995).

The rarely appreciated point is that the generalizable replication and application of results depends heavily on the existence of a portable and universally uniform observational framework. The inferences, judgments, and adjustments that can be made at the point of use by clinicians, teachers, managers, etc. provided with additive measures expressed in a substantively meaningful common metric far outstrip those that can be made using ordinal measures expressed in instrument- and sample-dependent scores.

These contrasts show that the confounding of statistics and measurement is a problem of vast significance that persists in spite of repeated efforts to clarify the distinction. In business, marketing, health care, and quality improvement circles, we find near-universal repetition of the mantra, "You manage what you measure," with very little or no attention paid to the quality of the numbers treated as measures. And so, we find ourselves stuck with so-called measurement systems where,

• instead of linear measures defined by a unit that remains constant across samples and instruments we saddle ourselves with nonlinear scores and percentages defined by units that vary in unknown ways across samples and instruments;

• instead of availing ourselves of the capacity to take missing data into account, we hobble ourselves with the need for complete data;

• instead of dramatically reducing data volume with no loss of information, we insist on constantly re-enacting the meaningless ritual of poring over indigestible masses of numbers;

• instead of calibrating instruments in an experimental test of the hypothesis that the intended construct is in fact structured in such a way as to make its mapping onto a number line meaningful, we assign numbers and make quantitative inferences with no idea as to whether they relate at all to anything real;

• instead of checking to see whether rating scales work as intended, with higher ratings consistently representing more of the variable, we make assumptions that may be contradicted by the order and spacing of the way rating scale categories actually work in practice;

• instead of defining a comprehensive framework for interpreting measures relative to a construct, we accept the narrow limits of frameworks defined by the local sample and items;

• instead of capitalizing on the practicality and convenience of theories capable of accurately predicting item calibrations and measures apart from data, we counterproductively define measurement empirically in terms of data analysis;

• instead of putting calibrated tools into the hands of front-line managers, service representatives, teachers and clinicians, we require them to submit to cumbersome data entry, analysis, and reporting processes that defeat the purpose of measurement by ensuring the information provided is obsolete by the time it gets back to the person who could act on it; and

• instead of setting up efficient systems for communicating meaningful measures in common languages with shared points of reference, we settle for inefficient systems for communicating meaningless scores in local incommensurable languages.

We ought not accept the factuality of data as the sole criterion of objectivity, with all theory and instruments constrained by and focused on the passing ephemera of individual data sets of local particularities. Properly defined and operationalized via a balanced interrelation of theory, data, and instrument, advanced measurement is not a mere mathematical exercise but offers a wealth of advantages and conveniences that cannot otherwise be obtained. We ignore its potentials at our peril.

William P. Fisher, Jr.
www.livingcapitalmetrics.com

References

Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath & S. H. Lovibond (Eds.), Mathematical and Theoretical Systems: Vol. 4 (pp. 7-16). North-Holland: Elsevier Science Publishers.

Cohen, J. (1994). The earth is round ($p < 0.05$). American Psychologist, 49, 997-1003.

Duncan, O. D. (1992, September). What if? Contemporary Sociology, 21(5), 667-668.

Falmagne, J.-C., & Narens, L. (1983). Scales and meaningfulness of quantitative laws. Synthese, 55, 287-325.

Hammersley, M. (1989). The dilemma of qualitative method: Herbert Blumer and the Chicago Tradition. New York: Routledge.

Wise, M. N. (Ed.). (1995). The values of precision. Princeton, New Jersey: Princeton University Press.

IOMW 2010: 15th International Objective Measurement Workshop

Wednesday, April 28th and Thursday, April 29, 2010
Immediately preceding AERA in Denver, Colorado, April 30 - May 4, 2010

University Memorial Center at the University of Colorado in Boulder, Colorado
near Denver and the AERA Conference Hotels.

The International Objective Measurement Workshop is a biennial conference devoted to the presentation discussion of topics germane to the theory and practice of measurement. All who are interested in methods of measurement in the context of education, psychology, sociology and medicine, are invited to participate as attendees and/or presenters.

Derek Briggs, conference chair. Nathan Dadey, the conference coordinator. www.iomw2010.net

Wednesday 4/28/2010

8:45 a.m. - 8:55 a.m. **Welcome - Derek Briggs**

9:00-10:45 Keynote Symposium

The BEAR Assessment System and Construct Development

1. Mark Wilson and Linda Morell - The BEAR Assessment System--State of the Art
2. Shih-Ying Yao, Eric Berson, Elizabeth Ayers, Sun-Joo Cho, and Mark Wilson - The Qualitative Inner-Loop of the BEAR Assessment System
3. In-Hee Choi, Yoonjeon Kim, Stephen Moore, and Mark Wilson - The Desired Results Developmental Profile: Validity Evidence for a Multidimensional, Multi-Age Instrument
4. Amy Dray, Yongsang Lee, Ronli Diakow, Nathaniel Brown, and Mark Wilson - The San Diego Striving Readers Assessment Project: Creating Measures of Reading Comprehension for Adolescent Readers

Discussant: Lorrie Shepard, University of Colorado at Boulder

11:00-12:15 Breakout Sessions - Symposia

Symposium 1: The Measurement of Learning Progressions in Science

1. Nathaniel Brown, S. Nagashima, A. Fu, M. Timms, and Mark Wilson - Analyzing Scientific Reasoning in Assessments Using Construct Modeling and the Evidence-Based Reasoning Framework
2. Linda Morell and Shih-Ying Yao - A Systematic Approach to Understand and Measure How Students Understand Science
3. Yongsang Lee, Jinnie Choi, and Karen Draney - The Carbon Cycle Learning Progression Project: Developing Empirically Based Learning Progression Variables of Science Literacy
4. Elizabeth Ayers, Robert Schwartz, and Shih-Ying Yao - Using Data and Analyses to Update Constructs

Symposium 2: Modeling Violations of the Local Independence Assumption

1. David Andrich - Conditional Estimates of Person Parameters in the Partial Credit Parameterization of the Rasch Model
2. Ou Zhang and Linjun Shen - Polytomous IRT or Testlet Model: An Evaluation of Scoring Models of Non-adaptive Assessment under Small Testlet Size Situation
3. Edward Wolfe and Aaron McVay - Rater Effects as a Function of Rater Training Context
4. Mary Garner - Using Paired Comparison Matrices to Analyze Connectivity and the Effects of Connectivity on Parameter Estimation

1:30-2:30 Breakout Sessions - Roundtables

Roundtable 1: Applications of Rasch Models: Facets, Dimensions & Structural Equations

1. Zairul Nor Deana Md Desa, Todd Little, and Matthew Gallagher - Multiple-Group MACS Models of High School Accountability Progress in Mathematics and English Language Arts

IOMW 2010 Theme:
Using Model Fit to Evaluate Hypotheses about Learning

The evaluation of fit in item response theory is often either not well understood or not given sufficient scrutiny. Casual rules of thumb for fit statistics are sometimes followed that may mask the presence of unusual - and revealing - response patterns. Furthermore, fit is often evaluated at the item level without giving equal scrutiny to fit at the person level. Beyond these problems, even when patterns of misfit have been identified, their diagnosis is often uncertain. For instance, the explanation for misfit could be found in the use of misaligned assessment items, indicating a faulty hypothesis of how growth along the construct develops or unusual characteristics in the makeup of the sample of students for whom empirical evidence has been collected.

One context in which rigorous evaluations of model fit are needed is the measurement of learning progressions. These are descriptions of increasingly sophisticated ways of thinking about or understanding a topic. Modelers are apt to declare success when the difference between what is predicted and what is observed seems rather small. Yet for assessment tasks that stem from a learning progression hypothesis, developers should not only be prepared to find considerable evidence of misfit, but they should embrace it and use it for revising and improving their instrumentation.

2. Karen Trimmer, Mohammed Quaddus, and Margot Wood - Sequential Use of Rasch Analysis and Structural Equation Modeling to Investigate Reasoned Risk-Taking in Decision-Making by School Principals
3. Stefanie Sebok and Peter MacMillan - Assessment of a Masters of Education Counseling Program Applicant Selection Process using a Multifaceted Partial Credit Rasch Model

Roundtable 2: Modeling Multiple Dimensions

1. Mark Moulton - Common History Equating: A New Method for Equating Multidimensional Formative Tests Without Common Persons or Items
2. Steffen Brandt - Increasing Unidimensional Measurement Precision Using a Rasch Subdimension Model
3. Andrew Maul - Examining the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) Using Multidimensional Rasch Modeling

Roundtable 3: Theory & Practice: Differential Item Functioning and Person Fit

1. Damian Birney, Elena Grigorenko, Steven Stemler, Jens Beckmann, and Robert Sternberg - An Investigation of the Theoretical Implications of Differential Item Functioning in the Evaluation of Educational Interventions
2. Aminah Perkins and George Engelhard, Jr. - Using Differential Person Functioning to Examine Student Performance on Mathematics Items
3. Sébastien Béland, David Magis, Gilles Raiche, and Nadine Talbot - Three Person-Fit Indexes with Estimated Ability Level: A Simulation Study

Roundtable 4: Measuring Teaching Practices

1. Brent Duckor -Measuring Assessment Literacy: An Item Response Approach
2. Sharon Osborn Popp, Audrey Beardsley, and Meredith Toth - Investigating the effectiveness of the Reformed Teaching Observation Protocol in a College of Education: A Many-Faceted Rasch Analysis
3. Nadine Talbot, Gilles Raiche, Sebastien Beland, Diane Leduc, Helene Meunier, and Valerie Djedje - Integration of Assessment of Learning into Teaching Practices: Validation of a Questionnaire on IPOD Touch

2:45-4:15 Keynote Symposium

Symposium: Historical and philosophical perspectives on measurement

1. George Engelhard, Jr . - An Overview of Historical and Philosophical Perspectives on Measurement
2. Nadia Behizadeh and George Engelhard, Jr. - Influences of Measurement and Writing Theories on Writing Assessment: An Historical Perspective
3. Laura Quaynor and George Engelhard, Jr. - Historical View of Measurement and Language Theories Within the Context of the Assessment of the Language Proficiency of English Language Learners
4. Margaret Keneman , Severine Piot, and George Engelhard, Jr. - Perennial Problems in the Assessment of Communicative Competence in a Second Language: A Rasch Perspective

Discussant: Finbarr Sloane, University of Colorado at Boulder

6:00 p.m. Happy Hour Reception, Downtown Boulder

Thursday 4/29/2010

9:00-11:00 Keynote Symposium

Using Residual Analysis in Understanding NCLB Results

1. Richard Smith - How Do Residuals Work?
2. Jeffery Thalberg - Using Residuals to Illustrate District Level Differences
3. Vincent Primoli - Using Residuals to Identify Strengths and Weaknesses in Text Books
4. Jeffery Thalberg - Using Residuals to Identify Sensitivity to Previous Performance
5. Vincent Primoli - Using Residuals to understand performance of LEP Students

Discussants: Greg Camilli & Ed Wiley, University of Colorado at Boulder

11:15-12:00 Breakout Sessions-Roundtables

Roundtable 1: Parameter Estimation Techniques

1. David Andrich - Metric Implications when Applying the Polytomous Rasch Model to Account for Response Dependence Between Item
2. Min Liu - Comparing Estimation for Rasch Mixture Models

Roundtable 2: Applications of the Partial Credit Model

1. Nathan Dadey and Magda Chia - Assessing Spelling Ability in English Language Learners
2. Diana Wilmot, Mark Wilson, Alan Schoenfeld, Danielle Champney, and William Zahner - Using Wright Maps to Understand Student Progress in College Readiness

Roundtable 3: Using Rasch Models to Measure Conceptual Understanding in Science

1. Heidi Iverson - Diagnostic Assessment of the Public Understanding of the Greenhouse Effect

2. Jacob Marszalek and Louis Odom - High School Biology Student Knowledge and Certainty about Diffusion and Osmosis Concepts

Roundtable 4: Applications of Item Response Models in Science

1. Colin Wallace, Edward E. Prather and Doug Duncan - Using the Rasch model in astronomy education research
2. Atar Burcu and Cobanoglu Aktan Derya - Latent Regression Analysis for TIMSS 2007 Physics Data for Turkish Students

Roundtable 5: Factor Structure Validation

1. Lihua Xu and John Hathcoat - Factor Structure Validation of the Inventory of School Motivation Among Chinese Students
2. Adaeze Nwaigwe - Can Learning Factors Analysis Method Measure Learning Progressions Across Different Subject Domains of Intelligent Tutoring Systems?

Roundtable 6: Instrument Validation & Standard Setting

1. Nu'man Saleh Al-Musawi - Development of a Scale to Measure Research Competence
2. Gregory Stone, Kristin Koskey, and Toni Sondergeld - Construct Definition in Angoff and Objective Standard Setting: Playing in a House of Cards without a Full Deck

12:30 - 1:00 **Poster Session**

1. Andrew Galpern - The Use of Flexible Scoring Procedures and Their Effects on Person and Item Fit Statistics
2. Man Hung - Investigating dimensionality – An uncompromised negotiation
3. Nathan Markward - "Genomic" Item Banks: Theory, Construction, and Implications for Expanding Applications of Personalized Medicine
4. Christian Spoden, Jens Fleischer and Detlev Leutner - Educational and psychological correlates of person fit in a large-scale mathematics assessment

1:15 - 2:30 Breakout Sessions - Symposia

Symposium 1: Communicating Results of the Rasch Model

1. Amy Subert and Matthew Gaertner - Measuring Voter Attitudes Toward Affirmative Action
2. Lois Lochhead and Peter MacMillan - Was the Sow's Ear ever a Silk Purse? Analysis of the Spinal Function Sort using the Rasch PCM
3. David Irribarra, Ronli Diakow, and Mark Wilson - Modified Wright Maps to Convey the Results of Ordered Partition Models
4. Adam Van Iwaarden - Using Rasch Fit Statistics in Small Sample Pilot Tests: Promise or Peril?

Symposium 2: Scale Interpretations in Construct Development

1. Kristin Koskey, Svetlana Beltyukova, Christine Fox, and Gregory Stone - Examining Quantitative and Qualitative Meaningfulness of Data Produced by Absolute Magnitude Estimation Scaling Using the Many Facets Rasch Model
2. Jackson Stenner and Mark Stone - Contrasting The Measurement of Temperature and Reading Ability: Implications For the Metrology - Psychometrics Distinction
3. Robert Cavanagh- Fitting Item and Person Data to a Scaling Model and to a Construct Model of One Dimension of Student Engagement
4. Anatoly Maslak, Nikolaus Bezruczko, Tatyana Anisimova, and Andrey Danilov - A Latent Trait Analysis of Higher Education Infrastructure in Russia

Symposium 3: Teacher Learning Progressions: Challenges and Opportunities for Articulating Growth in the Profession from a Rasch perspective

1. Mark Wilson - Issues in Defining and Assessing Teacher Learning Progressions
2. Brent Duckor - The Promise of Teacher Learning Progressions: The Case for Modeling Core Teacher Practice Strands in Pre-service Programs
3. Diana Wilmot - Using a Learning Progression to Understand Teachers' Developing Assessment Expertise

2:45 - 4:00 Panel Discussion: **Musings on the Future of Rasch Measurement**

In this session, a panel of scholars will respond and discuss questions related to future directions in the theory and application of Rasch Measurement. The idea here is to think big--what are the sorts of issues that we would want to see people addressing in research presented at the next IOMW in 2012? How can the principles of objective measurement be made relevant to policymakers and practitioners? There are many other questions that could be posed and we want you to pose them! **To submit questions use our online form: www.iomw2010.net/question.html**

Panelists:

* David Andrich, The University of Western Australia
* George Engelhard, Emory University
* Richard Smith, Data Recognition Corporation
* Mark Wilson, University of California, Berkeley

Moderator: Derek Briggs, University of Colorado at Boulder

Announcements from the Chair of the Rasch Measurement SIG

Greetings,

I am writing to provide several pieces of information relating to the Rasch SIG in advance of the upcoming AERA Annual Meeting.

First, please join me in congratulating and expressing appreciation to Michael Young and Kenneth Royal, who have been elected to two-year terms as the Rasch SIG Chair and Secretary/Treasurer, respectively.

Second, I am also happy to announce that a draft of text for Rasch SIG sponsored awards is available for review and comment by the SIG membership. I appreciate the work that Kelly Bradley and William Fisher have put into this draft over the last two years. As outgoing Chair, I will hand the future direction of this effort over to the incoming Chair, Michael Young. You can download and review a copy of the draft award guidelines here (http://www.edwolfe.net/files/Article_X_Awards_FisherBradley.docx). Please express your thanks to Kelly and William for their contributions to the SIG.

Third, I want to make you aware of the details for the Rasch SIG Business meeting. The meeting is scheduled for Saturday, May 1, from 6:15pm to 7:45pm in the Sheraton / Plaza Court 3. I will provide a brief State of the SIG address and hand the meeting over to the incoming Chair, Michael Young. Brent Duckor, of San Jose State University, will provide a presentation. Hors d'oeuvres (cheese and antipasto plates) and a cash bar will be provided.

Fourth, Secretary/Treasurer Timothy Muckle recently completed work on revisions of the Rasch SIG bylaws to make those bylaws conform to a template that all SIGs are now required by AERA to follow. The revisions are under review by AERA, and any changes that are made will be submitted to the SIG for approval sometime after the Annual Meeting.

Fifth, please remember to stop by the JAM Press booth in the Exhibit Hall while you are at AERA. Two new books should both be available by that time: Criterion Referenced Testing: Practice Analysis to Score Reporting Using Rasch Measurement and Advances in Rasch Measurement, Volume One. Thank you to Richard Smith, JAM's Editor, for providing this valuable venue for publishing Rasch relevant research.

Sixth, I encourage you to attend IOMW 2010, which will take place in Boulder, Colorado on April 28th and 29th. You can find the program for the conference on their website: http://www.iomw2010.net/. Thanks to Derek Briggs for Chairing the conference.

Finally, I am happy to provide you with the Rasch SIG's program at the 2010 Annual Meeting. Please join me in thanking Program Co-Chairs Diana Bernbaum (University of California, Berkeley) and Leigh Harrell (Virginia Tech) for their work in putting together this year's program. Also note that authors accepted for paper, roundtable and poster sessions should upload a copy of the final paper to be presented at the 2010 AERA Annual Meeting no later than April 9. Session chairs and discussants (if applicable) will review the uploaded papers through the system in order to prepare for the session. Instructions for uploading papers are shown below and are followed by the schedule of Rasch SIG presentations.

Upload and Online Papers Repository Instructions:

1. Login to the Online Program System using your ID# and password.
   (Click here https://www.aera.net/AALogin.aspx)
2. Click the link 'Read Reviews/Upload Final Draft/Online Repository'. You will see a list of all of your papers.

At the top of the page, click either 'yes' or 'no' to participate in the online paper repository.

Click the link 'Upload Final Draft' located to the right of the title for the paper you wish to upload. You will see an Upload Final Draft Screen.

Upload your paper and click the 'Upload and Continue' button at the bottom right of the screen.

If you need to upload a further revised paper before the April 9th deadline, just follow the instructions above. You cannot edit the paper once it has been uploaded. Uploading a new paper will remove the old paper.

Edward W. Wolfe, Ph.D.
Senior Research Scientist
Assessment & Information
Pearson
Email**:** ed.wolfe # pearson.com

2010 AERA Annual Meeting

Denver, Colorado

April 30 - May 4, 2010

Friday, April 30

2:15 p.m. to 3:45 p.m.

Roundtable Session 4: **Applications of Rasch Modeling for Student Learning** (Rasch Measurement SIG)
Building: Sheraton, Room: Grand Ballroom Section 2
Chair: Kelly D. Bradley, University of Kentucky

A Cross-Cultural Rasch Analysis of the Multidimensional Construct of Teachers' Feedback Practice, Eunlim Chi, Kyung Hee University, Korea; Jennifer Ann Quynn, University of Washington; Shin-Ping Tsai, University of Washington

Historical View of the Influences of Measurement and Writing Theories on the Practice of Writing Assessment, Nadia Behizadeh, Emory University; George Engelhard, Emory University

Rasch Analysis for School Engagement Survey, Pei-Hua Chen, University of Denver; Kathy E. Green, University of Denver

Roundtable Session 7: **Issues of Rasch Sampling and Test Equating** (Rasch Measurement SIG)
Building: Sheraton, Room: Grand Ballroom Section 2
Chair: Shungwon Ro, Prometric

A Practical Comparison of Test-Equating Methodologies to Measure College Readiness From 6th Through 12th Grade, Diana Bernbaum Wilmot, University of California - Berkeley

Comparing Routing Methods in the Multistage Test Based on the Partial Credit Model, Jiseon Kim, University of Texas - Austin; Hyewon Chung, John Jay College of Criminal Justice - CUNY; Barbara G. Dodd, University of Texas - Austin

Lessons From Evaluation and Equating Attempts With Self-Report Forms, Peter D. MacMillan, University of Northern British Columbia; Lois Lochhead, University of Northern British Columbia

4:05 p.m. to 5:35 p.m.

Expanded Applications of Item Response Theory
Building: Sheraton, Room: Governor's Square 17

A Multivariate, Multilevel Rasch Model for Measuring Teachers' Observed Instructional Practices Across Lessons and Time, Benjamin Kelcey, Wayne State University; Joanne F. Carlisle, University of Michigan; Geoffrey C. Phelps, University of Michigan; Daniel Berebitsky, University of Michigan; David J. Johnson, University of Michigan

Saturday, May 1

8:15 a.m. to 9:45 a.m.

Roundtable Session 9: **Use of IRT to Investigate Test Design, Mixed Formats, Common Item Characteristics, Linking Procedures and Item Location**
Building: Sheraton, Room: Grand Ballroom Section 2

A Random-Effect Rasch Model and a Random-Effect Response Time Model for Detecting Item Location Effect, Feiming Li, NBOME; Linjun Shen, National Board of Osteopathic Medical Examiners; Allan S. Cohen, University of Georgia

10:35 a.m. to 12:05 p.m.

Addressing Reliability and Validity Matters in Survey Research
Building: Sheraton, Room: Governor's Square 14

Reading Self-Efficacy Survey for Incarcerated Youth: Validity Measurement Based on Rasch Modeling, Weijia Ren, The Ohio State University; William Loadman, The Ohio State University; Raeal Moore, The Ohio State University; Jerome V. D'Agostino, The Ohio State University; Joy Edington, The Ohio State University; Anthony Vander Horst, The Ohio State University

Roundtable Session 12: **Research Methodology and Measurement**
Building: Sheraton, Room: Grand Ballroom Section 2

A Comparison of the Distributional Properties of Four Mean-Square Fit Indexes Utilized in Commercial Rasch Measurements Software, Mike McGill; Edward W. Wolfe, Pearson

12:25 p.m. to 1:55 p.m.

Roundtable Session 14: **Studies on Student Change and Effective Teacher Assessment Practices**
Building: Sheraton, Room: Grand Ballroom Section 2

Applying the Rasch Model to Measure Change in Student Performance Over Time, Jessica D. Cunningham, Western Carolina University; Kelly D. Bradley, University of Kentucky

Instrument and Scale Construction: Measurement in Survey Research Methods

Building: Sheraton, Room: Plaza Ballroom D

A Rasch Analysis on Collapsing Categories in Item Response Scales of Survey Questionnaires: Maybe It's Not One Size Fits All, Julie Grondin, Universite du Quebec - Rimouski; Jean-Guy Blais, University of Montreal

Rasch Rating Scale Analysis of the Attitude Toward Research Instrument, Elena C. Papanastasiou, University of Nicosia; Randall E. Schumacker, The University of Alabama

4:05 p.m. to 5:35 p.m.

Rasch Issues of Dimensionality, Scaling and Fit (Rasch Measurement SIG)

Building: Sheraton, Room: Plaza Court 6

Chair: Jon S. Twing, Pearson

Analysis of Unidimensionality Testing Procedures in Item Response Data via Principal Component Analysis of Residuals From the Rasch Model, Mike McGill; Edward W. Wolfe, Pearson

Comparison of Exposure Controls, Item Pool Characteristics, and Population Distribution: CAT With the Partial Credit Model, Hwa Young Lee, University of Texas - Austin; Barbara G. Dodd, University of Texas - Austin

Modeling the Rating Scores of Language Tests With Small Sample Sizes: A Comparison of Two Estimation Methods With the Many-Facet Rasch Model, Lixiong Gu, ETS; Guangming Ling, ETS; Frederick A. Cline, ETS

Performance of the Chi-Square Test for Detecting Unidimensionality in Applications of the Multidimensional Rasch Model, Leigh M. Harrell, Virginia Polytechnic Institute and State University; Edward W. Wolfe, Pearson

Simultaneous Modeling of Item and Person Dependence Using Multilevel Rasch Measurement Model, Hong Jiao, University of Maryland; Akihito Kamata, University of Oregon; Shudong Wang, Northwest Evaluation Association; Ying Jin, American Institutes for Research

The Impact of Test Model Change on Exam Passing Rate, Huijuan Meng, Pearson; Susan Steinkamp, Pearson

6:15 p.m. to 7:45 p.m.

Rasch Measurement SIG Business Meeting

Building/Room: Sheraton / Plaza Court 3

Chair: Edward W. Wolfe, Pearson

Presentation by Brent Duckor, San Jose State University

Abstract: Despite the promises of today's policy leaders and education entrepreneurs, the act of measuring teachers' effectiveness begs fundamental questions in psychometrics. First and foremost, what is the construct definition for "effectiveness?" What are its features, facets, and dimensions? How would we observe it? Which validation procedures are necessary to ensure the appropriate uses of score data? Researchers in the teacher effectiveness movement have failed to achieve consensus on the constructs, items designs, scoring strategies and appropriate measurement models that might be best suited to measuring teachers and calibrating observations. This talk develops a Rasch (1960) perspective from within the Constructing Measures (Wilson, 2005) framework to address the challenges and opportunities for modeling teacher learning progressions (Duckor, 2010).

Sunday, May 2

8:15 a.m. to 9:45 a.m.

Roundtable Session 22: **Use of Rasch Models in Instrument Validation** (Rasch Measurement SIG)

Building: Sheraton, Room: Grand Ballroom Section 2

Chair: Kristin L. K. Koskey, The University of Akron

Measuring Care: Using a Rasch Model to Verify Measurement of an Underlying Construct, Brandelyn L. Tosolt, Northern Kentucky University

Rasch Analysis of the Reasoning About Current Issues Instrument, Deborah Faust, University of Denver

The Relevance of Common Metrics: Investigating the Precision of Four Mindfulness Instruments, Sharon G. Solloway, Bloomsburg University of Pennsylvania

Using Rasch Modeling in the Analysis of a Qualitative Bully-Victimization Scale, Marybeth Lehto, University of Denver

10:35 a.m. to 12:05 p.m.

Roundtable Session 25: **Use of Rasch Model to Examine Measurement Error and Reliability** (Rasch Measurement SIG)

Building: Sheraton, Room: Grand Ballroom Section 2

Chair: Seock-Ho Kim, University of Georgia

An Index for Relative Error of Measurement Within Tolerance Intervals, Dimiter M. Dimitrov, George Mason University

Investigation of Rasch Measurement Precision Depending on the Number of Dichotomous Items, Anatoly Andreyevich Maslak, Slavyansk-on-Kuban St Pedagogical Institute

12:25 p.m. to 1:55 p.m.

An Expanded Look Into DIF

Building: Sheraton, Room: Governor's Square 14

Which Differential Item Functioning Indicator to Believe: The Rasch Model as a Method for Providing Evidence of Implementation Generalizability in a Randomized Control Trial, Andrew P. Swanlund, Learning Point Associates; Trisha Hinojosa, Learning Point Associates; Megan Brown, Learning Point Associates; Brenna O'Brien, Learning Point Associates; Kelly Ann Hallberg, Northwestern University

Roundtable Session 28: **Applications of the Rasch Model, Graded-Response Models, and Bayesian Techniques**

Building: Sheraton, Room: Grand Ballroom Section 2

Parameter Estimation of the Rasch Mixture Testlet Model Using the Marginal Maximum Likelihood Method, Hong Jiao, University of Maryland; Matthias Von Davier, ETS; Shudong Wang, Northwest Evaluation Association

2:15 p.m. to 3:45 p.m.

Student Learning in Mathematics and Design of Learning Tasks

Building: Colorado Convention Center, Room: Room 708

Evaluation of a Developmental Progression for Length Measurement Using the Rasch Model, Douglas H. Clements, University at Buffalo - SUNY; Julie Sarama, University at Buffalo - SUNY; Jeffrey E. Barrett, Illinois State University

Monday, May 3

8:15 a.m. to 9:45 a.m.

Using Item Response Theory to Investigate Various Challenges in Current Measurement Practice

Building: Sheraton, Room: Governor's Square 17

A Comparison of Global Fit Indexes in a Multidimensional Rasch Analysis of Polytomous Data, Leigh M. Harrell, Virginia Polytechnic Institute and State University; Edward W. Wolfe, Pearson

Roundtable Session 33: **Data Collection and Instrumentation Issues in Survey Research**

Building: Sheraton, Room: Grand Ballroom Section 2

Using the Principles of Survey Research and Rasch Measurement Results to Guide Refinement of a Survey Measuring Perceptions of Good Research, Kelly D. Bradley, University of Kentucky; Jessica D. Cunningham, Western Carolina University; Kenneth Royal, American Board of Family Medicine

12:25 p.m. to 1:55 p.m.

Studies on Rasch Conditions and Applications (Rasch Measurement SIG)

Building: Sheraton, Room: Plaza Court 6

Chair: Matthias Von Davier, ETS

Analysis of a College Placement Test in Mathematics Using the Rasch Measurement Model, Mary Garner, Kennesaw State University; Meghan Burke, Kennesaw State University

Crossing Person Response Functions: The Influences of Home Language, Gender, and Social Class on Mathematics Literacy in France, Germany, Hong Kong, and the United States, Aminah Perkins, Emory University; Laura Quaynor, Emory University; George Engelhard, Emory University

Evaluation of a Children's Mental Health Screening Instrument Using Rasch Rating Scale Methods, Christine DiStefano, University of South Carolina; Grant B. Morgan, University of South Carolina

Learning to Teach for Social Justice Beliefs: An International Construct Invariance Study, Larry H. Ludlow, Boston College; Michael O'Leary, St. Patrick's College; Fiona Ruth Ell, University of Auckland; Victor Bonilla, University of Puerto Rico; Marilyn Cochran-Smith, Boston College

Measuring Student Flow Experiences From a Classroom Engagement Perspective, Robert Frederick Cavanagh, Curtin University of Technology; Graham B. Dellar, Curtin University of Technology

Optimal Items for Assessing Parent Involvement in Early Childhood Identified Using the Rasch Model: Findings From the ECLS-K Data Set, Hui-Fang Chen Chen, Missouri State University

Tuesday, May 4

12:25 p.m. to 1:55 p.m.

Test Design and Item Selection Issues in Adaptive Testing

Building: Sheraton, Room: Governor's Square 17

Computerized Adaptive Testing for the Rasch Testlet Response Model With Ability-Based Guessing, Sheng-Yun Huang, The Hong Kong Institute of Education; Wen-Chung Wang, The Hong Kong Institute of Education

Rasch Model with an Error Term

Question: My referee insists that I write the Rasch model with an error term. How do I do that, and what is the error distribution?

Answer: When the Rasch dichotomous model is written with an error term it looks like this:

$$X_{ni} = P_{ni} \pm \sqrt{(P_{ni}*(1-P_{ni}))}$$

where X_{ni} is the scored response of person *n* to item *i*, and P_{ni} is the Rasch-model probability of a correct response, so that $P_{ni} = \exp(B_n-D_i) / (1+\exp(B_n-D_i))$, where B_n is the ability of person *n* and D_i is the difficulty of item *i*.

The distribution of each error term $\sqrt{(P_{ni}*(1-P_{ni}))}$ is binomial, because only two outcomes are possible for each observation, but when the error terms are accumulated across all the observations (as they are for estimation), the binomial errors approximate normality.

For a polytomous Rasch model,

$$X_{ni} = E_{ni} \pm \sqrt{(W_{ni})}$$

where E_{ni} is the Rasch-model expected value of the response and W_{ni} is the Rasch-model variance of the response around its expectation. The error distribution is multinomial, approximating normality across the dataset. See www.rasch.org/rmt/rmt34e.htm

ICOM 2010

Bethesda, Maryland
September 1-3, 2010
Call for Papers

The National Institutes of Health (NIH), Chestnut Health Systems, and the University of Illinois at Chicago are pleased to announce the upcoming *International Conference on Outcomes Measurement* (ICOM), www.esi-bethesda.com/icom2010
Call for Papers: all proposals must be submitted by **May 1st, 2010** to *smalla2 /at/ uic.edu* for consideration. Acceptances will be announced by June 1st, 2010.

ICOM's theme this year is "Modern Measurement: Focusing on Chronic Health Conditions." The goal is to advance the understanding of health screening and outcome issues in chronic health conditions, and applications of modern measurement models in these and related fields. Chronic illness outcomes applications and sessions concerning new theoretical developments in measurement will be included. We are excited to bring together clinicians, researchers and other health professionals in one setting to discuss these issues.

The conference will be held at the NIH Natcher Conference Center, near Washington, D.C., with workshop sessions to be held on Friday, September 2, 2010.

Shweta Malladi
ICOM Coordinator

Rasch-related Coming Events

Apr. 28-29, 2010, Wed.-Thur. IOMW 2010 International Objective Measurement Workshop, Boulder, CO, USA, www.iomw2010.net

Apr. 30 - May 4, 2010, Fri.-Tues. AERA Annual Meeting, Denver, CO, USA, http://www.aera.net

Apr. 30 - May 28, 2010, Fri.-Fri. Rasch - Core Topics (introductory) online course (M. Linacre, Winsteps), www.winsteps.com/courses.htm

May 12-14, 2010, Wed.-Fri. Introduction to Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

May 17-19, 2010, Mon.-Wed. Intermediate Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

June 7-9, 2010, Mon.-Wed. International Association for Computerized Adaptive Testing (IACAT) Conference, Arnhem, Netherlands, www.rcec.nl/iacat

June 13-16, 2010, Sun.-Wed. International conference on probabilistic models for measurement in education, psychology, social science and health, Copenhagen, Denmark, conference.cbs.dk/index.php/rasch

June 24 - July 4, 2010, Thur.-Sun. IX Summer School Measurement of Latent Variables (Rasch Measurement), Znanie, Russia, (A. Maslak), XIV Conference IX School

June 25 - July 23, 2010, Fri.-Fri. Many-Facet Rasch Measurement (intermediate) online course (M. Linacre, Facets), www.winsteps.com/courses.htm

June 29 - July 1, 2010, Tue.-Thur. PROMS-KL 2010 Pacific Rim Objective Measurement Symposium Kuala Lumpur, Malaysia, www.iiu.edu.my/proms/2010

July 26 - Nov. 20, 2010, Mon.-Sat. Online course: Introduction to Rasch Measurement of Modern Test Theory (Andrich, RUMM2030) www.education.uwa.edu.au/ppl/courses/introduction

Aug. 20 - Sept. 17, 2010, Fri.-Fri. Rasch - Core Topics (introductory) online course (M. Linacre, Winsteps), www.winsteps.com/courses.htm

Sept. 1-3, 2010, Wed.-Fri. ICOM 2010 International Conference on Outcomes Measurement, Bethesda MD, www.esi-bethesda.com/icom2010

Sept. 15-17, 2010, Wed.-Fri.Introduction to Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Sept. 20-22, 2010, Mon.-Wed. Intermediate Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Sept. 23-24, 2010, Thur.-Fri. Advanced Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Applications of Rasch Measurement in Science Education

Edited by **Xiufeng Liu**, State University of New York, Buffalo

and **William J. Boone**, Miami University (Ohio)

JAM Press is pleased to announce the new book, *Applications of Rasch Measurement in Science Education*, is available in soft cover and hard cover. More information is available at *JAM Press Books* www.jampress.org/jam-press.htm

The titles and authors of the twelve chapters are as follows:

Forward, *David F. Treagust*

1 Introduction to Rasch Measurement in Science Education, *Xiufeng Liu and William J. Boone*

2 Constructing a Quality Assessment through Rasch Techniques: The Process of Measurement, Feedback, Reflection and Change, *Kelly D. Bradley and Shannon O. Sampson*

3 The Geoscience Concept Inventory: Application of Rasch Analysis to Concept Inventory Development in Higher Education, *Julie C. Libarkin and Steven W. Anderson*

4 Tracking Cognitive Development with the Rasch Model: Empirical Evidence of Growth and Heterogeneity, *Lorna C. Endler and Trevor G. Bond*

5 Stage-like Patterns in the Development of Conceptions of Energy, *Theo L. Dawson*

6 Exploring Conceptual Understandings of Diffusion and Osmosis by Senior High School and Undergraduate University Science Students, *Debra Panizzon and Trevor Bond*

7 Mapping out Students' Matter Concept Development from Elementary to High School, *Xiufeng Liu*

8 Using Structured Item Response Theory Models to Analyze Content and Inquiry Reasoning Skills in BioKIDS, *Han Bao, Amelia Wenk Gotwals, Nancy Butler Songer, Robert J. Mislevy*

9 Assessing Students' Level of Knowledge and Analysing the Reasons for Learning Difficulties in Physics by Rasch Analysis, Alexander Kauert and Hans E. Fischer

10 Interpreting and Using Multidimensional Performance Data to Improve Learning in Science, *Cathleen A. Kennedy and Karen Draney*

11 Is the Paper and Pencil Assessment Mode Appropriate for Assessing the Learning Outcomes of Primary Science? Teachers' Attitudes, *Iris Lee and Guanzhong Luo*

12 Increasing Students' Interest: Informal Learning in Authentic Science Labs, *Katrin Engeln and Jürgen Rost*

4th UK Rasch User Group Meeting

Cambridge, UK

Report of 25th January 2010 meeting

Many thanks to Tim Croudace from The Psychometrics Centre, Cambridge, who hosted the successful fourth meeting of the user group. The programme included contributions from Evelina Galaczi from Cambridge ESOL, Gunter Maris from CITO in the Netherlands, Peter Tymms, John Little and John Tiffin from the University of Durham, Chris Wheadon from AQA, Sarah Muller from the University of Manchester and Tim Croudace from the University of Cambridge. Some of the slides provided by the participants are now available at the User Group website: www.rasch.org.uk/users/

See you all next year!

Chris Wheadon

Reviewing Rasch Papers

Request: "I have been asked to review a Rasch paper. What is the standard format for reporting Rasch analysis in publications?"

Response: www.jampress.org/guidelines.htm is a good place to start, but there is considerable variation in formats for Rasch papers. The most obvious difference is whether the paper focuses on the statistical properties of the numbers (*t*-tests, chi-squares, etc.) or the practical implications of the numbers (item maps, etc.).

The biggest problem seen in Rasch papers submitted for publication is that many authors do not understand their own analyses. Those authors parrot numbers from their output, but then they are unable to draw reasonable conclusions from those numbers. A consequence is "too many numbers, not enough meaning". Indications of this are Tables with unexplained or duplicative columns of numbers, and failure to use graphical Figures to communicate their Rasch findings effectively.

Criterion Referenced Testing:
Practice Analysis to Score Reporting Using Rasch Measurement

Edited by **Everett V. Smith, Jr.,** University of Illinois, Chicago,
and **Gregory E. Stone**, University of Toledo

JAM Press is pleased to announce the new book, *Criterion Referenced Testing: Practice Analysis to Score Reporting Using Rasch Measurement*, is available in soft cover and hard cover. More information is available at *JAM Press Books* www.jampress.org/jam-press.htm

The titles and authors of the 24 chapters are as follows:

1. Applications of Rasch Measurement to Job Analysis Data and the Translation into Content Weights. *Ning Wang*
2. Distractors with Information in Multiple Choice Items: A Rationale Based on the Rasch Model. *David Andrich and Irene Styles*
3. Item and Rater Analysis of Constructed Response Items via the Multi-Faceted Rasch Model. *Edward W. Wolfe*
4. Assessment of Differential Item Functioning. *Wen-Chung Wang*
5. Transitioning from Paper-and-Pencil to Computer-Based Testing: Examining Stability of the Rasch Latent Trait across Gender and Ethnicity. *Do-Hong Kim and Huynh Huynh*
6. Introduction to the Rasch Family of Standard Setting Methods. *Gregory Ethan Stone*
7. Psychometric Aspects of Item Mapping for Criterion-Referenced Interpretation and Bookmark Standard Setting. *Huynh Huynh*
8. Converging on the Tipping Point: A Diagnostic Methodology for Standard Setting. *John A. Stahl and Kirk A. Becker*
9. A Mapmark Method of Standard Setting as Implemented for the National Assessment Governing Board. *E. Matthew Schulz and Howard C. Mitzel*
10. Setting Passing Standards for Licensure and Certification Examinations: An Item Mapping Procedure. *Ning Wang*
11. Standard Setting with Dichotomous and Constructed Response Items: Some Rasch Model Approaches. *Robert G. MacCann*
12. Selecting Cut Scores with a Composite of Item Types: The Construct Mapping Procedure. Karen Draney and Mark Wilson
13. Objective Standard-Setting for Judge-Mediated Examinations. *Gregory Ethan Stone*
14. Evaluating the Judgments of Standard-Setting Panelists using Rasch Measurement Theory. *George Engelhard, Jr.*
15. The Mastery Level Judgment Consistency Rate of a Rasch Model Based Standard Setting Method for Classroom Achievement Tests. *Sun-Geun Baek and In Hee Choi*
16. Equating Designs and Procedures Used in Rasch Scaling. *Gary Skaggs and Edward W. Wolfe*
17. Tools for Measuring Academic Growth. *G. Gage Kingsbury, Martha McCall, and Carl Hauser*
18. Developing Examinations that use Equal Raw Scores for Cut Scores. *Andrew Swanlund and Everett Smith*
19. A Comparison between Robust z and 0.3-Logit Difference Procedures in Assessing Stability of Linking Items for the Rasch Model. *Huynh Huynh and Anita Rawls*
20. Equating of Multi-Facet Tests Across Administrations. *Mary Lunz and Surintorn Suanthong*
21. Computerized Adaptive Testing. *Ying Cheng and Leslie Keng*
22. The ISR: Intelligent Student Reports. *Ronald Mead*
23. Estimation of Decision Consistency Indices for Complex Assessments: Model Based Approaches. *Matthew Stearns and Richard M. Smith*
24. Deriving Proficiency Scales from Performance Indicators Using the Rasch Model. *Jean-Guy Blais, Michel D. Laurier, and Christian Rousseau*

When to stop removing items and persons?

Question: Each time I re-analyze my data after removing a misfitting item or person, another item or person misfits. When do I stop?

Answer: You need to ask yourself a practical question: Why am I removing items and persons? Is this your answer? "To improve the measurement of the persons!"

OK - then here is a strategy.

1. Estimate the person measures from the original analysis. Remove whatever persons and items you see to be really, really bad.

2. Estimate the person measures again. Cross-plot the person measures from 2. against those from 1. Are there any noticeable changes that matter to you?

No. Then the really, really bad wasn't so bad after all. Keep everything. Analysis 1 is what you want.

Yes. Now remove the really bad persons and items.

3. Estimate the person measures again. Cross-plot the person measures from 3. against those from 2. Are there any noticeable changes that matter to you?

No. Then the really bad wasn't so bad after all. Analysis 2 is what you want.

Yes. Now remove the somewhat bad.

4. Estimate the person measures again. Cross-plot the person measures from 4. against those from 3. Are there any noticeable changes that matter to you?

No. Then the somewhat bad wasn't so bad after all. Analysis 3 is what you want. Stop here.

Yes. ..... (and so on)

If you must report all the persons or items, anchor (fix) the good measures at their final good values. Reinstate everything that you need to. Reanalyze and do your final reporting. The good items and person are anchored so that the bad stuff will not distort their measures.

You may discover that the conventional fit criteria are much too strict and are strongly influenced by the misfit of only a few observations.

Rasch Measurement Transactions
14525 SW Millikan Way #27615, Beaverton OR 97005
www.rasch.org/rmt
Editor: John Michael Linacre

SIG Chair: Ed Wolfe
Secretary: Timothy Muckle
Program Chairs:
Diana Wilmot and Leigh Harrell
SIG website: www.raschsig.org

Removing 50% of the Items

Question: I must remove nearly 50% of the items to meet the recommended fit criteria. What is wrong with my Rasch analysis?

Reply: Where does misfit end? Imagine you are cleaning a window.
First you see big dirty areas - so you remove the dirt.
Then you see smudges - so you wipe off the smudges.
Then you see scratches - so you polish out the scratches.
Then you see distortions - so you grind out the distortions.
Then you notice discolorations - so you .....

At each stage of a fit analysis, the data are changed so the definition of "good fit" changes. We stop when the measures are "good enough", in the same way as we stop when the window is "clean enough".

One way to verify that the measures are "good enough" is to output the person measures at each stage, and then to cross-plot the person measures for each stage against those of the previous stage. When the difference between the later stage and the earlier stage is too small to matter, or the person measures for a later stage are less meaningful than for an earlier stage (often because there are too few items remaining), then the measures for the earlier stage are "good enough".

Doing this "good enough" procedure, we may discover that the original misfit in the data is too small to have any meaningful impact on the measures. So that no editing of the data is needed.

Journal of Applied Measurement
Volume 11, Number 1. Spring 2010

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

The Ministry of Education and Science of the Russian Federation
Slavyansk-on-Kuban State Pedagogical Institute
Slavyan Chapter of Institute for Objective Measurement
Laboratory for Objective Measurement

XIV International Workshop Theory and Practice of Measurement of Latent Variables
IX Summer School Measurement of Latent Variables (Rasch Measurement)

June 24 – July 4, 2010 on the coast of Black sea – in the resort "Znanie" ("Knowledge")

The Summer School Program: Lecturer: **DSc., Prof. Anatoli A. Maslak**

1. Introduction to the Measurement of Latent Variables
2. Traditional Test Theory: Merits and Drawbacks
3. Measurements Based on Rasch Model
4. Analysis of Test and Test Items Quality
5. Computer Training on Estimation of Rasch Model Parameters
6. Investigation of Rasch Measurement Precision
7. Rasch Measurement in Education, Psychology, Sociology, Public Health
8. Rasch Measurement in Russia and Abroad
9. Round Table: "Role of the Fulbright Program in Promoting Modern Technologies in Social Sciences"

Full announcement at: XIV_Conference_IX_School

Which Hypothesis is the Null Hypothesis?

The choice of null hypothesis depends on the focus of our investigation. In dimensionality analyses, we could say: "My null hypothesis is that these data are multidimensional. We need strong evidence to reject multidimensionality." Or "My null hypothesis is that these data are unidimensional. We need strong evidence to reject unidimensionality". Both hypotheses are valid, but we do need to make sure that our audience is clear about our premise.

This situation is encountered throughout statistical analysis. Which hypothesis is the null hypothesis? It is our choice depending on how we conceptualize the problem to be solved. For instance, a correlation of 0.0 is the null hypothesis when we are looking for a correlation between variables, but a correlation of 1.0 is the null hypothesis when we are looking for a lack of collinearity between variables.

Online Measurement Programs at UIC

The University of Illinois at Chicago is offering two online programs in research methodology. The first program is an eight course *MEd in Measurement, Evaluation, Statistics, and Assessment* (MESA).

education.uic.edu/mesaonline-med/

The second program is an *Educational Research Methodology (ERM) Certificate*, which consists of a minimum of any three courses offered in the MESA online curriculum. Those interested in taking a course without entering a program can enroll as an Extramural Student.

education.uic.edu/erm/

Everett Smith, Ph.D.

Rasch in Genetics

From the Elsevier website:

Paper: Rasch-based high-dimensionality data reduction and class prediction with applications to microarray gene expression data.

Authors: Andrej Kastrin and Borut Peterlin, Institute of Medical Genetics, University Medical Centre Ljubljana, Slovenia.

Journal: Expert Systems with Applications (online 2010)

Abstract: Class prediction is an important application of microarray gene-expression data analysis. The high-dimensionality of microarray data, where number of genes (variables) is very large compared to the number of samples (observations), makes the application of many prediction techniques (e.g., logistic regression, discriminant analysis) difficult. An efficient way to solve this problem is by using dimension-reduction statistical techniques. **Increasingly used in psychology-related applications, Rasch model (RM) provides an appealing framework for handling high-dimensional microarray data.** In this paper, we study the potential of RM-based modeling in dimensionality reduction with binarized microarray gene expression data and investigate its prediction accuracy in the context of class prediction using linear discriminant analysis. Two different publicly available microarray data sets are used to illustrate a general framework of the approach. Performance of the proposed method is assessed by re-randomization scheme using principal component analysis (PCA) as a benchmark method. Our results show that RM-based dimension reduction is as effective as PCA-based dimension reduction. The method is general and can be applied to the other high-dimensional data problems.