**Wright's Rasch Regression**

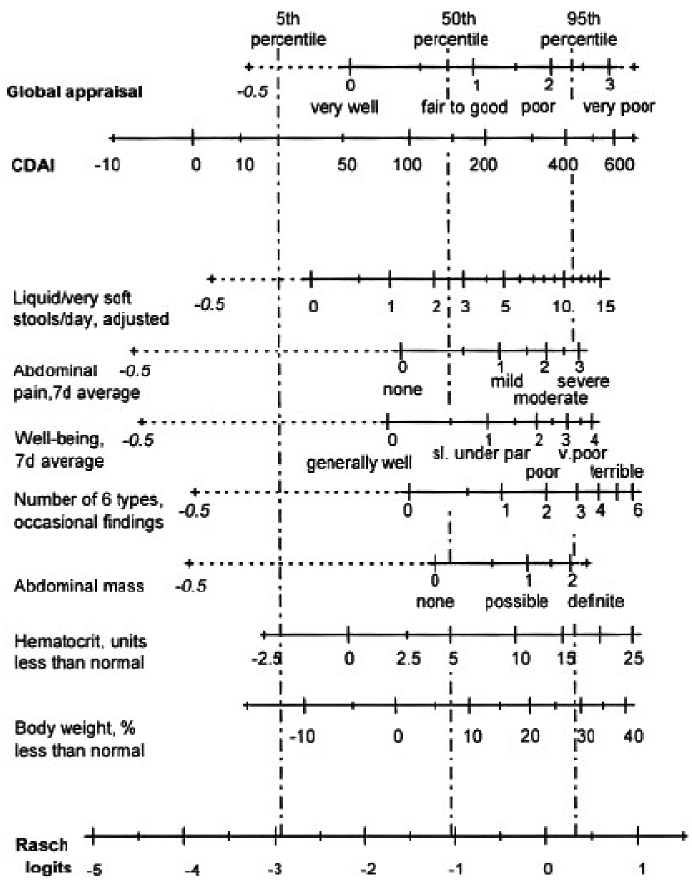
```
YARDSTICK DEFINED BY SCORES ON THE THREE TESTS
  20    30[34]40   [50]   60    70    80
-----+-----+-|---+-----|-----+-----+-----+----
       :     | 4      :|  6  : 8 : 9  : 11   :
          : 24)23    : 20 18(17 : 13 :10: 7: 3
 :   6  :   9| : 12 :16|:19: 21 : 25   :   28
-----+-----+-|---+-----|-----+-----+-----+----
  20    30[34]40   [50]   60    70    80
-----+-----+-|---+-----|-----+-----+-----+----
       NO   :    ?    :   YES
```

# RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 22 No. 3 Winter 2008 ISSN 1051-0796

# A Construct Map that Ben Wright Would Relish




This construct map of the "Crohn's Disease Activity Index" (CDAI) is Figure 11 in William R. Best, "A Rasch Model of the Crohn's Disease Activity Index (CDAI): Equivalent Levels of Ranked Attribute and Continuous Variable Scales", Chapter 5 in "Crohn's Disease: Etiology, Pathogenesis and Interventions", Jack N. Cadwallader (Ed.), Nova Publishers, 2008.

The map shows the relationship between the highly non-linear clinical indicators and the "Rasch logits" linear variable of "degree of sickness attributed to Crohn's Disease". Equal degrees of sickness align vertically on each of the clinical variables.

In a clinical setting, as soon as even one indicator variable has been observed, an estimate of the overall Rasch logit measure can be made. Observing more indicator variables refines the Rasch measure, but expensive, intrusive or time-consuming indicators are only required when high precision is needed for clinical decisions.

Changes in status across time can be tracked accurately in equal-interval Rasch logits, in the same way that weight or temperature can be tracked. The clinician's judgment is not misled by the non-linearity of the indicator variables.

Ben Wright discusses these maps further at "Rasch Regression - My Recipe", RMT 14:3, p. 758-9 www.rasch.org/rmt/rmt143u.htm

*William Best summarizes the technical aspects of his analysis for RMT:*

A partial credit Rasch model was developed from the original 112 patients' data used in the 1970s to develop the CDAI. For this model, continuous variables were condensed into variables of 6 ranked attributes each. Relationship of each of the nine component CDAI variables to Rasch logits is explored through scattergrams with jittered coincident points. Logic plus examination of these graphs indicate that best-fit scales should be smooth, monotonic, and with decreasing logit steps for equal interval variable steps. An origin-shifted logarithmic equation, encompassing observed ranges of a variable (X) and logit (Y), fits these specifications well; $Y = a + b \log_e(X + c)$.

An objective in this effort is to juxtapose nine component CDAI scales and a single Rasch logit scale, so a viewer may appreciate equivalencies within observed ranges. This requires that the relationship between any two variables be the same in both directions. In conventional regression they are not. The curve that best describes equivalency between two scales is that passing through

**Table of Contents**

the same percentile values of those scales. Percentile is calculated as p=100(3i-1)/(3n+1), where i is rank and n is total.

Good fit to the origin-shifted logarithmic model is generally possible using three points, low, middle, and high. 5-50-95 percentiles were chosen for this sample, resulting in about 6 values in each tail. One cannot usually infer equal-interval relationships between values in ranked-attribute variables. However, assuming a continuous, unbounded distribution permits data manipulations not otherwise possible.

A simple model used here assigns integers to categories, 0 through k, with range, -0.5 to k+0.5. Zero implies a finding absent or at normal levels. Ties at any integer level, i, are then distributed uniformly between i-0.5 and i+0.5, rendering each percentile unique. Whether this approach is credible in any case depends on the face validity of resultant formulations.

An alternate approach for continuous variables is Deming regression, also producing a single relationship independent of direction. Ideally, one needs the error variance of each variable and computations are complex. However, assuming proportionate error is often intuitively appropriate and greatly simplifies computation.

I would welcome published reviews of this chapter.

*William R. Best*

## Introduction to Rasch Measurement of Modern Test Theory EDU8636

The Graduate School of Education,
The University of Western Australia (UWA)
Professor David Andrich and Dr Ida Marais

## EXTERNAL STUDY/ONLINE UNIT

### 20 July - 14 November 2009 (Semester 2)

*Introduction to Rasch Measurement of Modern Test Theory* is a graduate unit of study introducing Rasch measurement: www.rasch-analysis.com

The unit is available in external study mode. This means that the unit can be studied from anywhere in the world. A discussion site will operate via the *Learning Management System* as part of the unit of study. Participants also obtain a copy of the RUMM software for the duration of the course.

The Unit can be taken towards a degree in any university, but students need to check in advance that their university will give credit.

The Unit can also be taken as a professional development course without participants being enrolled in a university degree.

**Please contact Ms Natalie Carmody for more details:**
**natalie.carmody ~at~ uwa.edu.au** or +61 8 488 2308

David Andrich, BSc., MEd. (UWA), PhD (Chic), FASSA
Chapple Chair, Graduate School of Education
The University of Western Australia
david.andrich ~at~ uwa.edu.au
www.matildabayclub.net/

## To: Rasch Measurement SIG Members From: the SIG Chair

As we approach the annual meeting of the American Educational Research Association, I wanted to thank you sincerely for your continued involvement in the Rasch Special Interest Group. Your support over the years has resulted in our maintaining a strong Rasch SIG membership – 125 members!

Before the meeting, I wanted to write to you to make you aware of the following:

AERA is moving to a calendar year basis for membership, meaning that annual memberships expire on 12/31 for all members. Please mark you calendars so as not to miss this important deadline. If your membership is due to expire, please go to the AERA web site to log in and renew. Be sure to renew your Rasch SIG membership during this process!

AERA is also moving to a unified elections system, so the next Rasch SIG elections (to be held in fall of 2009) will be held through that system. This will move up the timeline for securing nominations. Rasch SIG Chair Ed Wolfe will contact the membership in September of 2009 regarding the nominations time-frame. Please consider running for either the Chair or Secretary positions.

Two of our active members, William Fisher and Kelly Bradley, are working as an *ad hoc* committee to draft a proposal for two Rasch SIG awards. We hope to send you information regarding these awards as the date of the 2009 Annual Meeting approaches.

The Rasch SIG web site ( www.raschsig.org ) is in need of a face lift. A number of changes and updates are needed in the near future. We need one of our members to serve as a volunteer webmaster to make these updates and to take the lead role in managing the website. If you are willing and able, please contact SIG Chair Ed Wolfe at *ed.wolfe ~at~ pearson.com* .

Two of our members, Diana Bernbaum and Dimiter Dimitrov, served as program co-chairs for the 2009 meeting. That process ran smoothly, and our thanks go to them for providing this service to the Rasch SIG membership. If you are interested in serving as a program co-chair for the 2010 annual meeting, please contact Ed Wolfe, the Rasch SIG Chair.

I hope you plan to attend the Annual Meeting of the AERA in San Diego in April. If you do, please be sure to join us at the *Rasch SIG business meeting,* which will be held on Tuesday, April 14th from 6:15 – 7:45 PM in the *Omni San Diego, Room Gaslamp 1.*

*Tim Muckle, SIG Chair*

# Repeated Measure Designs and Rasch

"Subjects in repeated-measures designs provide more than one score" (Introduction to Analysis of Variance, J. R. Turner, Sage, 2001). This motivates well-intentioned statisticians to proclaim that "since the repeated measures cannot be independent, the Rasch model is not appropriate." But this statement can be paradoxical.

If the analysis or repeated-measures is based on raw scores, then the analysis is treating the raw scores as though they represent all the relevant information in the underlying observations. In other words, the raw scores are the "sufficient statistics" for the underlying observations. Further, if a raw score is to represent one quantitative ability, then it must not matter which pattern of observations on a given set of items generated that raw score. If we apply these considerations to a set of responses, we discover that they require the raw scores to fit the Rasch model. See "Rasch Model from Raw Scores as Sufficient Statistics", RMT 3:2, p. 62, www.rasch.org/rmt/rmt32e.htm

If the reviewers doubt that the Rasch model is applicable, they are also doubting that the raw score is an accurate summary of the observations.

This suggests that the statisticians have their logic backwards. If we have raw scores from a repeated-measures design, we need to submit their underlying observations to Rasch analysis in order to discover whether the raw scores are locally independent enough that they can form the basis for valid statistical analysis.

What if the Rasch analysis does indicate that the raw scores are defective? One approach is to select a subset of the observations that do fit the Rasch model. Use this subset to generate definitive item measures and definitive rating-scale structures (Rasch-Andrich thresholds). Then perform an analysis of the entire dataset with the items and thresholds anchored (fixed) at their definitive values. Then the dependency among the repeated measures will not distort the Rasch person measures.

Overly predictable data, appearing as overfit to a Rasch model, are rarely a problem because the data become redundant (not needed), but the measures correspond to the data. Underfit to the model is a problem because the measures do not accurately correspond to the data.

*(Suggested by Tsair-Wei Chien, Chi-Mei Medical Center, Tainan, Taiwan.)*

## An Ode to Ben Wright

by Carl Granger, 2009

There was a fellow named Benjamin Wright.
Psychology, Physics and Mathematics were his delight.
For believers in Rasch, he taught with great might.
For non-believers, he would give them a great fright.
Although he never would fight,
he always enjoyed being Wright!

Composed in honor of Ben Wright's 83rd birthday, Monday, March 30, 2009. Ben Wright will speak at the celebratory Symposium: "Improving Efficiency in Health Outcome Measurement", Chicago, Saturday, March 28, 2009.
www.ric.org/research/centers/cror/projects/rrtc/data/T3.aspx

## Pacific Rim Objective Measurement Symposium (PROMS 2009)

### July 28-30, Tues.-Thurs., 2009
### Hong Kong

Please consider joining other Rasch measurement enthusiasts in Hong Kong for the *Pacific Rim Objective Measurement Symposium* – July 28, 29 & 30, 2009. Our concept of the Pacific Rim is very broad – we already have proposals from Serbia, South Africa, Germany, & the UK, as well as from SE Asia.

Here is the link to the PROMS web-site:
www.promshk.org
and a note of welcome from the PROMS co-chairs, Profs Mok and Wang of The Hong Kong Institute of Education:

It is our great pleasure to welcome you to the Pacific Rim Objective Measurement Symposium 2009 (PROMS 2009) Hong Kong, to be held in Hong Kong on 28, 29 & 30 July 2009, with workshops scheduled on 27 July 2009.

The Centre for Assessment Research and Development (CARD) and the Department of Educational Psychology, Counselling and Learning Needs (EPCL) of The Hong Kong Institute of Education are delighted to have the opportunity to co-host the scientific event for PROMS 2009 HK. We count on your participation to make this event successful.

PROMS 2009 HK focuses on recent advances of objective measurement. It aims to provide an international forum for discourse on the latest research in using Rasch measurement as a tool for scientific progress. You are invited to join with a panel of distinguished researchers and other practitioners to share their expertise and your experiences in objective measurement.

Hong Kong is the host city for PROMS 2009 HK. We promise you memorable experiences in this vibrant international city of the Asia Pacific where East meets West.

We look forward to welcoming you to Hong Kong.

Magdalena Mo Ching MOK & Wen Chung WANG,
Conference Co-Chairs, PROMS 2009 HK

*You might think that the future of Rasch measurement is in the US or Europe – you might just be wrong! ;-)*

Trevor Bond
PROMS HK 2009 Honorary Adviser

# Common-Item Equating with Different Test Discriminations

If some items are the same in two different tests, then these may be chosen as "common items". When we perform a Rasch-analysis of the each test separately, we will obtain an estimate of the item difficulty of each common item on each test.

We expect that a scatterplot of the pairs of item estimates for the common items will have a trend line effectively parallel to the identity line. Figure 1 shows this situation in which the 12 common items lie along a trend line (dotted) parallel to the identity line (arrows). 0.0 logits for Form A (x-axis) corresponds to 1.1 logits for Form B (y-axis). To equate Form B with Form A, we subtract 1.1 logits from the Form B item and person measures.

But the trend line may not be parallel to the identity line. What do we do? Figure 2 shows a situation in which the trend line is not a line (arrow) parallel to the identity line. But the arrow falls with in the 95%-confidence error bands (black curves). We cannot reject the hypothesis that the trend line is statistically parallel to the identity line.

Under these circumstances, there are two ways forward.
1. We can choose the Form which we consider to be more definitive. Then obtain its item difficulties and Rasch-Andrich thresholds. Use these values as anchor (fixed) values in an analysis of the other Form. This puts the measures of the second form into the measurement frame-of-reference of the first form.
or 2. We can combine the data from the two Forms into one analysis (concurrent equating) in which all the items and persons are measured in the same frame-of-reference.

Figure 3 shows the situation in which the trend line (dotted line) is obviously and statistically not parallel to the identity line (arrow). Now we have tests of different discrimination, and a "Fahrenheit-to-Celsius" (F-C) equating situation, similar to that encountered with temperature conversions. The "Fahrenheit" measures on Form B needs to be rescale as "Celsius" measures compatible with Form A. An equation for doing this is:

Form B rescaled as Form A = (mean of Form A common items) + (Form B measure - mean of Form B common items) * S.D. of Form A common items / S.D. of Form B common items.

If we have two tests with common items that we know to be functionally different (such as from different test publishers, or in different presentation formats) then we expect to do F-C equating.

If we have two tests with common items that are supposed to be the same (such as alternate test forms, or pre-test and post-test forms), then we are reluctant to do F-C equating. We usually decide which form is the "correct" form (or combine the two forms) and use it as the basis for the equating.

*(Suggested by Mg. Andrés Burga León, Ministerio de Educación del Perú)*

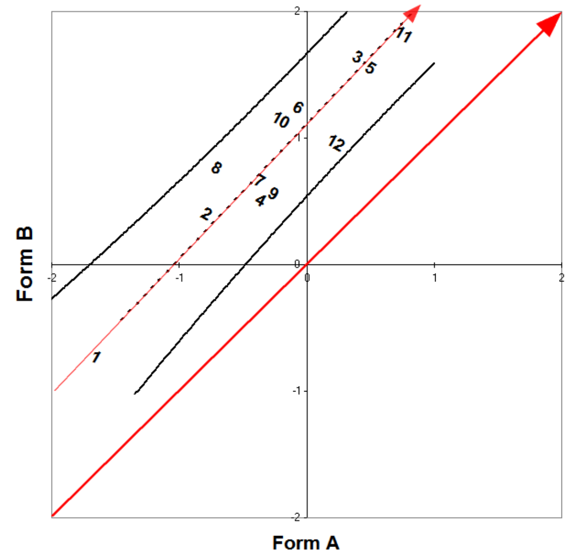


Figure 1. Exact common-item equating

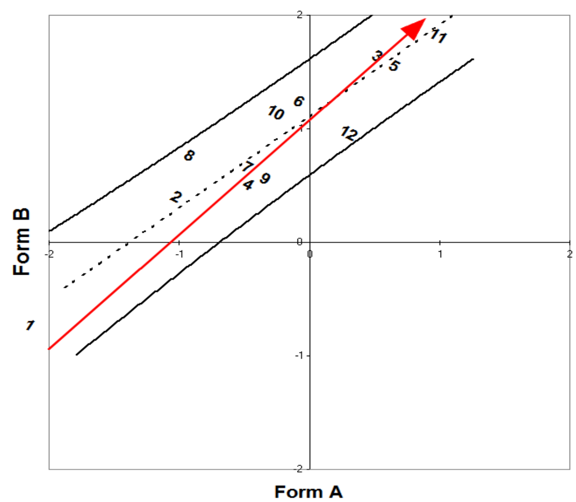


Figure 2. Statistically-equivalent equating

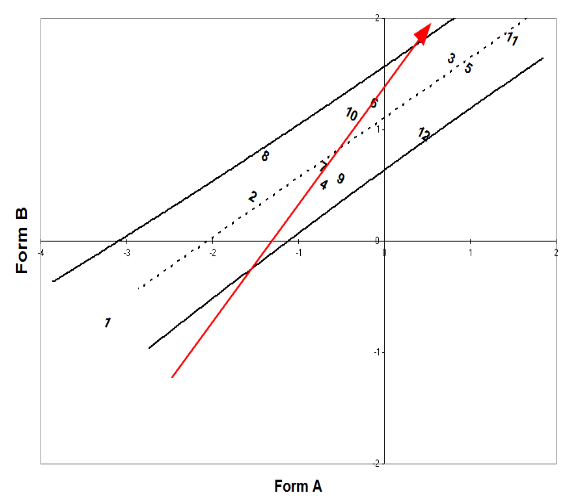


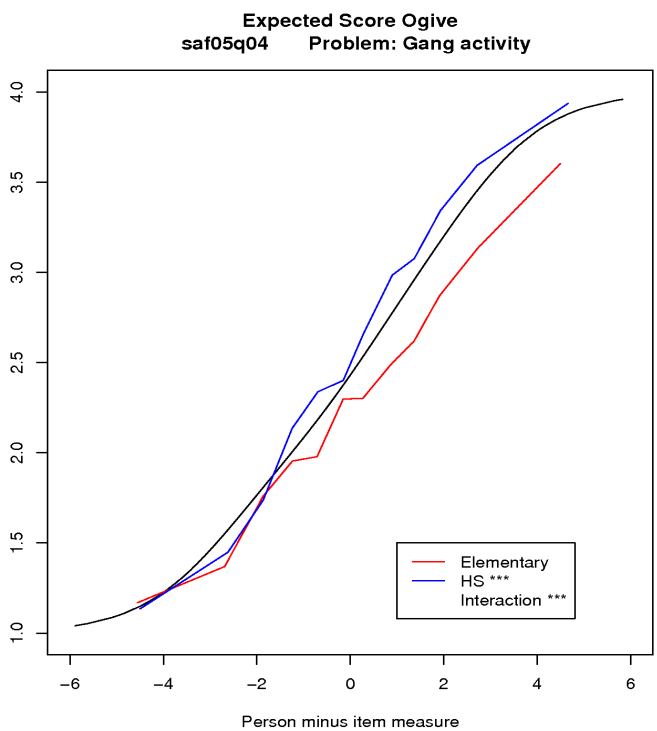
Figure 3. Fahrenheit-to-Celsius Equating

# Advanced DIF Analysis with Winsteps

Winsteps users: do you suffer from RUMM2020 envy? Do you wish you could enjoy the nifty non-uniform DIF detection features that only RUMM2020 users get? Do you wish you could break down the expected responses to polytomous items, and plot them by measure level the way they do it in RUMM2020? RUMM2020 has some very cool features, but Winsteps users do not need to left out in the cold. Thanks to the freely available, open-source data analysis and graphics system, R, Winsteps users can do non-uniform DIF analysis and graphing at least as well as you can with RUMM2020.

R is an open-source implementation of the S language, which was developed at Bell Laboratories. Versions are available for free download for Unix, Windows and MacOS X from www.r-project.org . R comes with a wide range of built-in functions. But since it is a very powerful programming language, extensions can be written to do practically anything. There are at least 1600 contributed packages already available to do almost anything you can think of. At least three packages do Rasch analysis. If you can't find a package already written that will do what you need, you can always write your own. That is what I did in this case.

I have been using the standardized item difficulty difference method of DIF detection for many years. This is fine for dichotomous items, and for uniform DIF in polytomous items, but I began to suspect that groups of survey respondents were responding differently depending on their position on the scale. Since I do not have RUMM, I decided on a DIY approach.




The R program I wrote accomplishes the task like this:

1) Calculates the probability of responding in each of the categories.

2) Calculates the expected response for each person using the category probabilities.

3) Calculates standardized residuals of actual responses from expected responses by group for ranges along the scale.

4) Does 2-way ANOVA on the residuals by group and scale range, with an interaction term.

5) Plots the expected score ogive and the empirical score curves for each group and displays the results of the ANOVA in the graph.

The Figure is an example of such a plot. This item is part of a scale measuring teachers' perceptions of crime and disorder in the school. Teachers are asked to state how much of a problem they perceived gang activity to be.

The black line is the expected score ogive; the blue line is the empirical responses from high school teachers; the red line, the responses from elementary school teachers. At the very low end of the scale, there is very little difference between the responses of high school and elementary teachers. However, at the high end of the scale, high school teachers report that gang activity is more of a problem than elementary school teachers do. The asterisks by "High School" in the legend indicate that the difference between high school and elementary teachers, overall, is very significant. In addition, the asterisks by "Interaction" indicate that the difference across the range of the scale is also significant.

The R program for DIF graphing is available at
http://home.uchicago.edu/~lupp/all-DIF.R

It is heavily commented, but will need modification because the group selection mechanism is unique to each person's situation.

R Statistics - www.r-project.org
RUMM2020 - www.rummlab.com.au
Winsteps - www.winsteps.com

*Stuart Luppescu*
*Consortium on Chicago School Research*
*University of Chicago*

**Rasch Measurement Transactions**
P.O. Box 811322, Chicago IL 60681-1322
www.rasch.org/rmt
Editor: John Michael Linacre

## An Application of the Mixed Rasch Model

"The mixed Rasch model (Rost, 1990) is an extension of the more traditional Rasch model, basically making *diagnostic* use of differential behavior difficulties of similarly constrained or facilitated – latent – groups of people. The mixed Rasch model – in contrast to the latent class model – also allows for different performance levels within latent groups of people. The mixed Rasch model can be used to identify different groups based on a set of predictors (test items), such as different ecological behaviors. Using the groups' differential behavior difficulties *diagnostically* means finding out about potential behavioral consequences of certain contextual factors. By applying the mixed Rasch model confirmatorily to differential behavioral consequences of real environments, valid contextual information and method bias can be successfully discriminated. We see the confirmatory application of the mixed Rasch model as a powerful *diagnostic* tool for disclosing situational factors responsible for facilitating and constraining certain performances beyond people's volitional control."

*excepted from:*

Kaiser, F.G., Keller, C. (2001). Disclosing situational constraints to ecological behavior: A confirmatory application of the mixed Rasch model. *European Journal of Psychological Assessment, 17,* 212-221.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14,* 271–282.

## Journal of Applied Measurement
## Volume 9, Number 4. Winter 2008



*Richard M. Smith, Editor*

JAM web site: www.jampress.org

## Rasch-related Coming Events

Feb. 27 - March 27, 2009, Fri.-Fri. Rasch - Further Topics online course (M. Linacre, Winsteps), www.statistics.com/ourcourses/raschfurther

March 11-13, 2009, Wed.-Fri. Introduction to Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

March 15, 2009, Sun. Online Programs in Research Methodology (Fall 2009, registration deadline), www.rasch.org/onlineuic.htm

March 28, 2009, Sat. Symposium: Improving Efficiency in Health Outcome Measurement, Chicago www.ric.org/research/centers/cror/projects/rrtc/data/T3.aspx

April 13-17, 2009, Mon.-Fri. AERA Annual Meeting, San Diego, CA, USA, www.aera.net

May 1-29, 2009, Fri.-Fri. Many-Facet Rasch Measurement online course (M. Linacre, Facets), www.statistics.com/ourcourses/facets

May 13-15, 2009, Wed.-Fri. Introduction to Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

May 18-20, 2009, Mon.-Wed. Intermediate Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

June 2-3, 2009, Tues.-Wed. 2009 GMAC Invitational Conference on Computerized Adaptive Testing, Minneapolis, www.gmac.com/CATConference

June 26 - July 24, 2009, Fri.-Fri. Rasch - Core Topics online course (M. Linacre, Winsteps), www.statistics.com/ourcourses/rasch1

June 29 - July 2, 2009, Mon.-Thur. 9th International Symposium on Measurement Technology and Intelligent Instruments, Russia, www.tdisie.nsc.ru/ismtii2009

July 20 - Nov. 14, 2009, Mon.-Sat. Introduction to Rasch Measurement of Modern Test Theory online course (D. Andrich, I. Marais, RUMM), www.rasch.org/andrich2009.htm

July 28-30, 2009, Tues.-Thur. PROMS HK 2009 Pacific Rim Objective Measurement Symposium, Hong Kong www.promshk.org

Sept. 6-11, 2009, Sun.-Fri. IMEKO XIX World Congress: Fundamental and Applied Metrology, Portugal www.imeko.org

April 30 - May 4, 2010, Fri.-Tues. AERA Annual Meeting, Denver, CO, USA, www.aera.net

June 14-15, 2010, Mon.-Wed. International Conference on Probabilistic Models for Measurement, Copenhagen, Denmark. www.rasch2010.cbs.dk