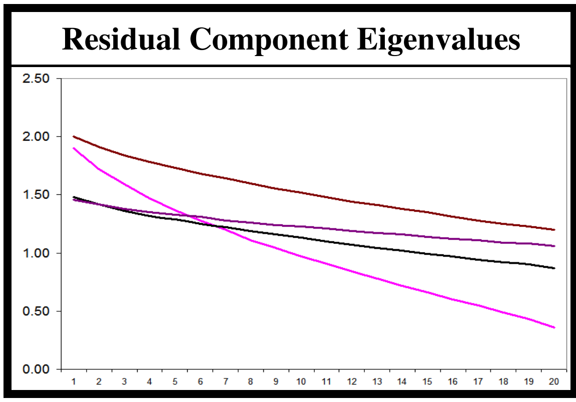


RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 19 No. 1 | Summer 2005 | ISSN 1051-0796

Dichotomous & Polytomous Category Information

Huynh & Mayer (2003) present some useful findings regarding the statistical information provided by ordered categories. Their work suggests a further device for identifying the location of categories on the latent variable.

Let θ be a location on the latent variable relative to the item difficulty (defined as the point on the latent variable at which the highest and lowest categories are equally probable to be observed). $P_k(\theta)$ is the probability of observing category *k* at location θ. Then, the expected value of the observation at θ is E(θ) where

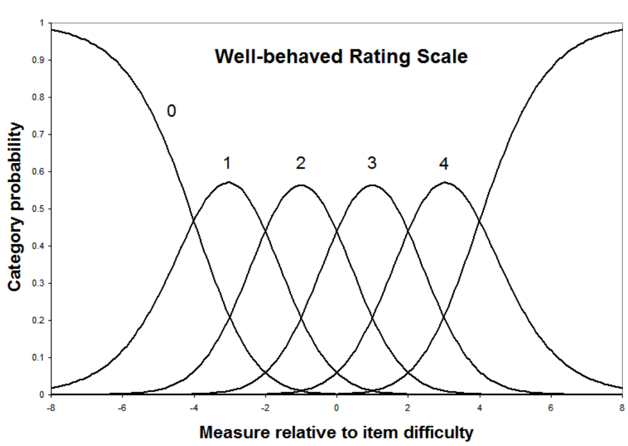
$$E(\theta) = \sum_{k=0}^{m} k P_k(\theta)$$

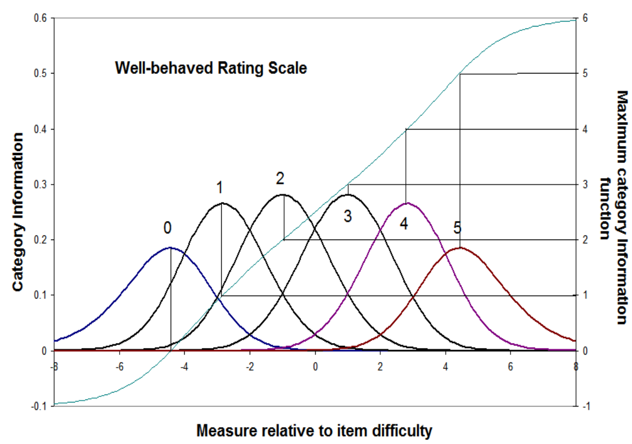
where *k*=0,m are the m+1 categories of the dichotomy or polytomy. E(θ) is the model ICC (item characteristic curve).

Let I(θ) be the Fisher information in an item at that location. Then I(θ), the item information function, corresponds to the slope of the item characteristic curve, the item's model variance, at that point.

$$I(\theta) = \sum_{k=0}^{m} (k - E(\theta))^2 P_k(\theta)$$

and let S(θ) be the skewness of the item at θ, so that





$$S(\theta) = \sum_{k=0}^{m} (k - E(\theta))^3 P_k(\theta)$$

Then $I(\theta)P_k(\theta)$ is the information that can be attributed to category *k* at θ. For every category, its information at the extremes of the latent variable is asymptotically zero. At some point or points along the latent variable the information peaks. At a maximum, the differential of the category information is zero, i.e., where:

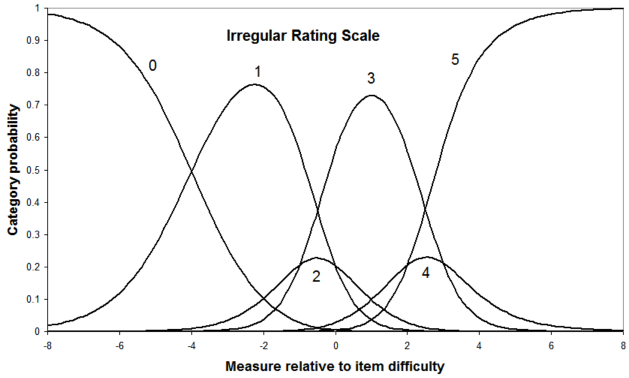
$$\frac{dI(\theta)P_k(\theta)}{d\theta} = 0$$

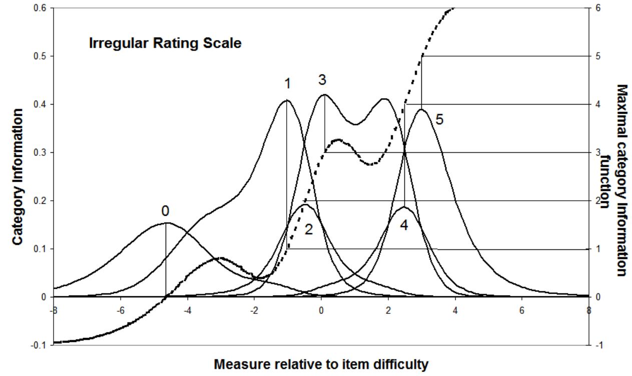
In general,

$$\frac{dP_k(\theta)}{d\theta} = P_k(\theta)(k - E(\theta))$$

$$\frac{dE(\theta)}{d\theta} = I(\theta)$$

Table of Contents





$$\frac{dI(\theta)}{d\theta} = S(\theta)$$

$$\frac{dI(\theta)P_k(\theta)}{d\theta} = P_k(\theta)(S(\theta) + I(\theta)(k - E(\theta)))$$

so that category information maxima (and minima) occur where

$$k = E(\theta) - S(\theta)/I(\theta)$$

Contrast this with the parallel expression for where the category probability maxima occur:

$$k = E(\theta)$$

An advantage of the category information approach is that it identifies locations for the maximum information of the extreme categories, while such locations do not exist for the maximum probabilities of extreme categories.

Here are the category probabilities, category information and maximum information curve for a well-behaved polytomous item. The maximum information for the extreme categories occurs, in this example, at ±4.4 logits where the extreme categories have a probability of 0.6. The measures are more central, and the probabilities are lower than other approaches suggest. For instance, the measure corresponding to a probability of 0.75 for the extreme categories is ±5.1 logits.

For a less well-behaved rating scale with uneven category probabilities, the information function is more complex, and can have multiple maxima for one category. In this irregular example, the probability of the lowest category at the location of maximum information. -4.6 logits is only .64. For the highest category, it is at 3.9 logits where the category probability is .83.

John Michael Linacre

Huynh H. & Meyer P.L. (2003) Maximum information approach to scale description for affective measures based on the Rasch model. *Journal of Applied Measurement, 4, 2,* 1010-110.

External Study/Online Unit, 25 July – 31 October 2005

Introduction To Rasch Measurement And Traditional Test Theory

Unit Coordinators: Professor David Andrich and Dr Ida Marais

www.education.murdoch.edu.au/educ_RaschCourse2005.html

The Unit Of Study - Background

In the Australian Semester 2, 2005 (July 26 to November 26), a graduate unit of study introducing Rasch measurement is available in the *external study* mode. This mode of study means that the unit can be studied from anywhere in the world. A discussion group will operate for online interaction as part of the unit of study. Students enrolled obtain (i) a set of lecture materials, which includes hard copy of all of the lectures, (ii) details of the assignments you will be required to submit, (iii) the necessary reading materials, and (iv) the Study Guide setting out the steps you will need to follow to successfully complete the unit. This unit has been presented in the same period every year from 2000. In each of 2002, 2003, and 2004, over 50 people from many parts of the world took the opportunity to enroll. Because of the success of the previous presentations, the course is being offered again this year.

Examples of positive responses to the Unit in the past:

"This is by far one of the best courses on measurement theory I have ever enrolled in!"

"Despite it being a distance course, I learned a great deal."

"Both unit materials and assignments allowed me to learn the essential aspects of the subject."

"The lecture materials were well organized, logical, and easy to follow."

Features Of The Unit

1) it begins from first principles,
2) exercises at the end of each lecture consolidate the ideas,
3) it introduces the Guttman structure as a lead into both traditional test theory and Rasch measurement,
4) it reviews elementary traditional test theory in a way that it relates to the Rasch models,
5) it reviews the necessary elementary statistics,
6) it studies the dichotomous model and the model for ordered response categories,
7) it studies model fit, including differential item functioning,
8) it involves discussion group which permits you to interact with other students in the class, and it provides a full version of the interactive, Windows program RUMM2020 for analyzing data. (The use of the program is available throughout the unit)

The RUMM program is a very easy to use interactive program that permits learning many features of the Rasch measurement model by working around the program's menus – for example the effects of rescoring any item, deleting items, studying alternatives in distracters, assessing differential item functioning, automatic linking of different sets of items, effects of deleting samples or individuals, taking account of missing data, and so on. To enhance understanding all of the information is available both graphically and statistically, including item characteristic curves, person item maps, etc

Who Should Enroll

The unit is suitable for people from many social research backgrounds, but four in particular have been seen to gain most benefit from their enrolment.

(i) Professionals engaged in assessment and measurement of performance and attitude, who know traditional test theory and are interested in learning the principles of modern test theory and Rasch measurement in particular;

(ii) People in education, psychology, health care, health sciences who are concerned with outcome measurement;

(iii) People who have become familiar with Rasch measurement and item response theory through professional exposure, but would like to consolidate their understanding of its first principles;

(iv) Students, including graduate students in doctoral programs who are involved in higher degree studies and require knowledge and evidence of studying educational and psychological measurement, in particular introduction to traditional and modern test theory.

Topics Covered

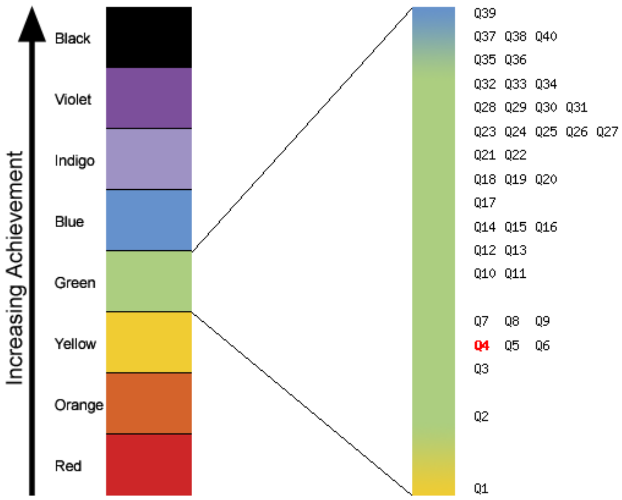
1	Review of measurement and statistics in education and social science
2	Reliability and validity
3	Formalization of traditional reliability
4	Calculation of reliability
5	The Rasch model for dichotomous responses: The simplest latent trait model:
6	Separation of person and item parameters
7	The significance of total scores
8	Estimating person ability and item difficulty
9	Fit of the data to the model (1)
10	The Rasch model for ordered response categories: Analysis of partial credit or rated items
11	Fit of the data to the model (2). Differential Item Functioning (DIF)
12	(a) A relationship between the reliability of traditional test theory and Rasch latent trait theory (b) Linking and equating using the Rasch model

David Andrich

iAchieve online assessment services

"Where You're at Report"

Colored bands indicate performance levels. Test questions are positioned by difficulty. The question in **bold red** was answered incorrectly. ***www.iachieve.com.au***



Journal of Applied Measurement Volume 6, Number 2. Summer 2005

Estimating item parameters in the Rasch model in the presence of null categories. G*uanzhong Luo and David Andrich*

Effects of item redundancy on Rasch item and person estimates. *Everett V, Smith, Jr.*

Measuring progress towards smoking cessation. *Melinda F. Davis, Lee B. Sechrest, and Dan Shapiro*

Daredevil barnstorming to the tipping point: new aspirations for the human sciences. *William P. Fisher, Jr.*

Comparing Rasch analyses probability estimates to sensitivity, specificity and likelihood ratios when examining the utility of medical diagnostic tests. *Daniel Cipriani, Christine Fox, Sadik Khuder, and Nancy Boudreau*

A rank-ordering method for equating tests by expert judgment. *Tom Bramley*

Using the Rasch model to validate stages of understanding the energy concept. *Xiufeng Liu and Sarah Conrad*

Book Review – Designing and using tools for educational assessment, Madhabi Chatterji. *Joanna M. Kulikiwich*

Book Review – Making measures, Benjamin D. Wright and Mark H. Stone. *Edward W. Wolfe*

Richard M. Smith, Editor
Journal of Applied Measurement
P.O. Box 1283, Maple Grove, MN 55311
JAM web site: www.jampress.org

Erling Andersen: *In Memoriam*

Professor Erling B. Andersen deceased a few days before his 65 year anniversary September 18, 2004. He planned to retire by the end of 2004, but sudden illness made an unexpected end of these plans.

As the first elected chairman of the *Danish Society for Theoretical Statistics*, he was a key person in creating a platform for young graduated statisticians, which. by the end of the 1960's, were the first result of changes in the field of mathematics at the University of Copenhagen. (At a meeting the *Society* in February, 2005, Svend Kreiner presented "The Danish Rasch tradition: An appreciative examination of Erling B. Andersen's contributions to the theory of Rasch models.")

Erling considered himself as the first true student of G. Rasch and was employed for many years in positions related to Georg Rasch as a person and, in particular, related to theoretical developments of what later are denoted as Rasch Models. In a way this relationship never changed as a student-teacher relationship, although Erling soon contributed by independent research to various aspects of analyses of the Rasch Model. Maybe Erling too often stressed the technical issues of the models rather than the philosophy behind the models? His thesis "Conditional Inference and Models for Measuring" of 1973 contained the later widely used conditional testing of the Rasch Model. We remember G. Rasch as the first official opponent when Erling was defending his thesis, taking the chair and addressing for a very, very long time just the interpretation of the title "Models for Measuring" – should it be "for measuring" or, rather "for measurements"?

Along with a number of international papers directed towards the analysis of the general Rasch Model for more than two response categories, Erling's books "Discrete Statistical Models with Social Science Applications" (1980) and "The Statistical Analysis of categorical Data" (1994) became standard text books for many students in the field of analysis by latent variables relating to, for example, generalizations of the Rasch Models.

Erling changed over the last couple of years from a person always cheerful and interested in matters outside his own world to be more introvert and to be not very easy to socialize with. Many of us, his colleagues and friends, are proud to have known Erling and highly appreciate what he did for the international spread of knowledge about the Rasch Models. We are at the same time sad that he didn't get a better opportunity to harvest the fruits of his life works as Professor Emeritus of G. Rasch's former chair at the Institute of Statistics, University of Copenhagen.

Prof. Peter Allerup
Department of Educational Psychology
The Danish University of Education
Copenhagen, Denmark

The Partial Credit Model and the One-Item Rating Scale Model

At least one aspect of Rasch measurement continues to perplex analysts and paper reviewers. Are Masters' Partial Credit Model and Andrich's Rating Scale Model variants of the same polytomous model or different models?

The Andrich (1978) Rating Scale Model conceptualizes all items on an instrument to share the same m+1 ordered-category rating scale:

$$P_{nix} = \frac{e^{x(B_n - D_i) - \sum_{k=0}^{x} F_k}}{\sum_{j=0}^{m} e^{j(B_n - D_i) - \sum_{k=0}^{j} F_k}}$$

with the usual parameterization, where x=0,m and F_0=0 or any convenient constant, also $\sum_{k=0}^{m} F_k$ =0.

The Masters' (1982) Partial Credit Model conceptualizes each item to exhibit a unique rating scale structure of m_i+1 ordered categories.:

$$P_{nix} = \frac{e^{xB_n - \sum_{k=0}^{x} D_{ik}}}{\sum_{j=0}^{m_i} e^{jB_n - \sum_{k=0}^{j} D_{ik}}}$$

where x=0,m_i and F_{i0}=0 or any convenient constant.

In many survey instruments, subsets of items share rating scales, some items have unique rating scales, and some items are dichotomies. The generalization of the Andrich Rating Scale Model to groups of items encompasses all these:

$$P_{ngix} = \frac{e^{x(B_n - D_{gi}) - \sum_{k=0}^{x} F_{gk}}}{\sum_{j=0}^{m_g} e^{j(B_n - D_{gi}) - \sum_{k=0}^{j} F_{gk}}}$$

where *g* indicates the group of items (sharing the same rating scale structure) to which item *i* belongs.

But what if every group contains only one item? Then *g* has the same meaning as *i,* and this model becomes:

$$P_{nix} = \frac{e^{x(B_n - D_i) - \sum_{k=0}^{x} F_{ik}}}{\sum_{j=0}^{m_i} e^{j(B_n - D_i) - \sum_{k=0}^{j} F_{ik}}}$$

Now it appears that there are two different Rasch models for the identical situation: the "Partial Credit" and the "One-Item Rating Scale". What is the relationship between them? Let us take the Partial Credit model and reparameterize D_{ik} as $b_i + \tau_{ik}$ where $b_i = \sum_{k=0}^{m_i} D_{ik} \Big/ m_i$

Then,

$$\sum_{k=0}^{m_i} D_{ik} = \sum_{k=0}^{m_i}\left(b_i + \tau_{ik}\right) = \sum_{k=0}^{m_i}\left(b_i\right) + \sum_{k=0}^{m_i}\left(\tau_{ik}\right)$$

$$= m_i b_i + \sum_{k=0}^{m_i}\left(\tau_{ik}\right) = \sum_{k=0}^{m_i} D_{ik} + \sum_{k=0}^{m_i}\left(\tau_{ik}\right)$$

Therefore, $\sum_{k=0}^{m_i}\left(\tau_{ik}\right) = 0$ which is the same constraint as in the one-item rating scale model.

Thus, the difference between the two models is reduces to parameterization. The "Partial Credit" D_{ik} is identical to the "one-item Rating Scale" $D_i + F_{ik}$ as constrained by

$$D_i = \sum_{k=0}^{m_i} D_{ik} \Big/ m_i \quad \text{and} \quad \sum_{k=0}^{m_i}\left(F_{ik}\right) = 0$$

Consequences of this equivalence include the definition of an overall "item difficulty" for a Partial Credit item as D_i, and also any theoretical properties or practical implications obtained for one model can be carried directly over to the other.

D_i has a convenient interpretation: it is the location (i.e., person measure) on the latent variable at which the highest and lowest category are equally probable. To confirm this, let B_n be the ability of person *n* with equal probability of being observed in the lowest and highest categories of item *i* of difficulty D_i:

$$P_{ni0} = P_{nim}$$

$$\frac{e^{F_0}}{\sum_{j=0}^{m} e^{j(B_n - D_i) - \sum_{k=0}^{j} F_k}} = \frac{e^{m(B_n - D_i) - \sum_{k=0}^{m} F_k}}{\sum_{j=0}^{m} e^{j(B_n - D_i) - \sum_{k=0}^{j} F_k}}$$

$$e^{F_0} = e^{m(B_n - D_i) - \sum_{k=0}^{m} F_k}$$

$$F_0 = m\left(B_n - D_i\right) - \sum_{k=0}^{m} F_k$$

$$m\left(B_n - D_i\right) = \sum_{k=1}^{m} F_k = 0$$

$$B_n = D_i$$

Thus item difficulty for Andrich's Rating Scale model and Masters' Partial Credit model can have the same definition. The models are equivalent.

John Michael Linacre

Andrich D. (1978) A rating scale formulation for ordered response categories. Psychometrika, 43, 561-573.

Masters G.N. (1982) A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Book Review: Applying The Rasch Model

Applying the Rasch Model, Fundamental Measurement in the Human Sciences, by Trevor G. Bond and Christine M. Fox (2001). Mahwah, NJ: Lawrence Erlbaum Associates, Inc, 255 pages. ISBN 0-8058-4252-7.

Applying the Rasch Model is a book that should be embraced by the research community as a foundation to the basic properties and principles of Rasch measurement. There is no question that the authors have achieved their goal of presenting an accessible overview that would not require a sophisticated statistical background. This book would draw even the most novice researcher into the topic. It is a first-rate selection for the practicing researcher who wants a tutorial on the Rasch model, for the faculty looking to find an introductory text on Rasch measurement, or for those who desire to extend their knowledge of various Rasch models and their applications. Bond and Fox have done an excellent job of taking a difficult task and making it understandable and useful to the research community.

This is the first paragraph of a Book Review by Kelly D. Bradley, University of Kentucky, which appeared in Organizational Research Methods, 2005, 8, 2, 249-250.
http://orm.sagepub.com/content/vol8/issue2/

A Rasch Citation Anomaly

According to Google-Scholar, there are **388** citations of Georg Rasch's book, "**Probabilistic** Models", and **84** citations of "**Probabalistic** Models".

Let's use our spelling-checkers!

MOMS
Midwest Objective Measurement Seminar
Friday, May 20, 2005, 9:00 – 4:00

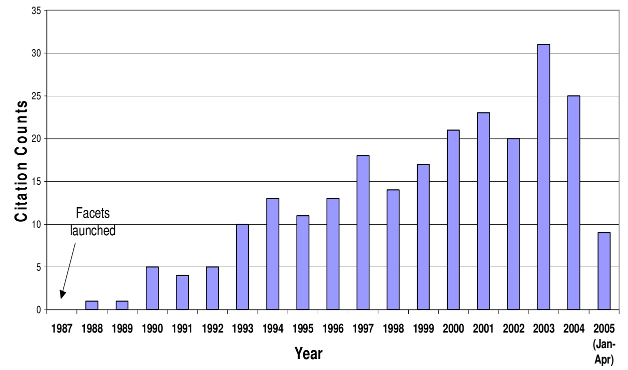
School of Public Health
University of Illinois at Chicago
1603 West Taylor Street
Chicago IL 60612-4394

The purpose of the MOMS is to allow students, teachers and practitioners to present research in progress or results of research studies, as well as, provide a forum for discussion of measurement issues. The seminar provides the opportunity for those interested in measurement to get together for serious discussion and social interaction. Please send your proposals to me. The proposal should be one paragraph that explains the purpose of the study, data, and results summary. Be sure to include your name and affiliation. Email proposals to me at mlunz-at-measurementresearch.com

I encourage all of you to participate.
Mary E. Lunz, www.measurementresearch.com

Sponsored by the Institute for Objective Measurement and the University of Illinois at Chicago

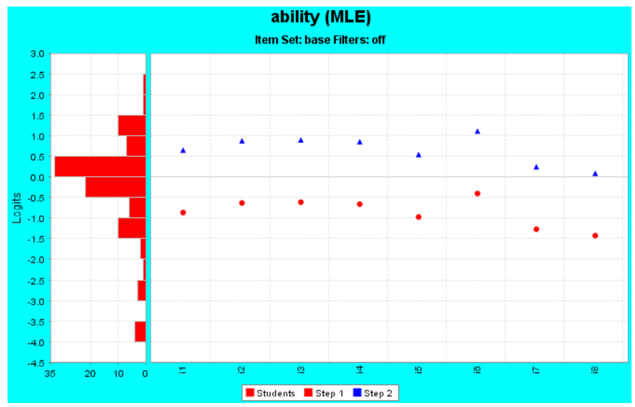
Citations to Facets



Scholarly citations by year to Facets Rasch software (Linacre, 1987) and synonyms. The 241 citations are listed at: www.winsteps.com/facetman/references.htm

---

GradeMap is Progressing Quickly!



Now downloadable: Version 4.1 of GradeMap, developed in association with Mark Wilson's book "Constructing Measures" (Erlbaum).

FAMILIA DE PROGRAMAS KALT
San Salvador, El Salvador

Introductory Course on Rasch Analysis
conducted by Agustin Tristan (in Spanish)
July 20, 2005

Evaluation Conference
organized by the Institute of Evaluation and Advanced Engineering, Mexico
July 21-22, 2005

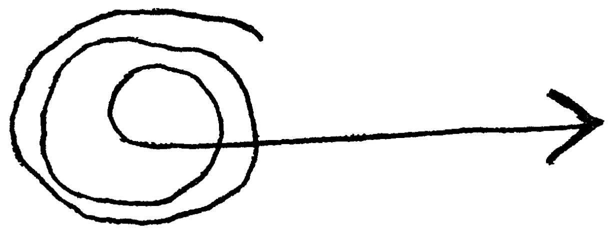
Organized Sight-Seeing: July 23, 2005

Participation fee: $175 includes materials, a CD of the presentations and lunch. Hotel rate $50 single, $55 double includes buffet breakfast.
Air transport is affordable, with flights from Houston, Mexico City, San Jose, etc. People are friendly and there are no security problems.

web page (in Spanish and English)
www.ieesa-kalt.com

SOCIAL *spiral* to SCIENCE *arrow*

Ben Wright



1. DESIGN a Conversation — Initiate an **arrow**

Aim a **construct** at a **respondent**
Spiral into meaning
Fish out **replications**
of a **within respondent** line of inquiry

Define the **arrow**
Focus **questions** into **items**
Craft **response** alternatives
Pilot in person

2. SAMPLE the Conversation — Pursue the **arrow**

Replicate among respondents
Pilot small groups

3. FOCUS Responses — Clarify the **arrow**

Rank alternatives
Eliminate superseded responses
Pivot items

4. BUILD Measures — Document the **arrow**

Confirm **Construct**
Items: map, separation, fit

Illustrate **Application**
Persons: map, separation, fit

Benjamin D. Wright (2001) p. 71 in "Adventures in Questionnaire Design: Poetics, Posters and Provocations", B.D. Wright, Marci M. Enos, Matthew Enos, and J.M. Linacre, Chicago: MESA Press.

Probabilities direct the conduct of the wise man.
(Probabilia ... sapientis vita regeretur.)
Cicero, De Natura Deorum, i, 5, 12. (45 B.C.?)

But to us, probability is the very guide of life.
Bishop Joseph Butler, The Analogy of Religion (1736)

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

Critical Eigenvalue Sizes in Standardized Residual Principal Components Analysis

A principal components analysis of Rasch residuals, i.e., of observed responses minus their expectations, is used in Wright (1996) to investigate whether or not there is more than one component explaining the structure of respondent data. Wright postulates that, if the data are unidimensional, then components in the residuals will be at the noise level. Wright uses logit residuals. Linacre (1998) argues in favor of residuals standardized by their model standard deviation. These have the form of random normal deviates and will be adopted here.

The idea of retaining components that are above noise level is common practice in psychometrics. The Cattell (1966) scree test and the Kaiser (1960) rule are the most often used procedures to determine the number of components. They are both based on inspection of the correlation matrix eigenvalues. Cattell's recommendation is to retain only those components above the point of inflection on a plot of eigenvalues ordered by diminishing size. Kaiser (1960) recommends that only eigenvalues at least equal to one are retained. One is the average size of the eigenvalues in a full decomposition.

Smith and Miao (1994, p. 321) observe many components with eigenvalues greater than one in four simulations of unidimensional observational data. In their simulations, the first component corresponds to the Rasch dimension. The eigenvalue of the second component, the largest component in the random noise, never exceeds 1.40, suggesting that 1.40 is a threshold value for randomness.

Humphreys and Montanelli (1975) argue that the Kaiser rule is only true for very large correlation matrices. They propose that criterion eigenvalue thresholds be estimated by simulation studies based on random data formed into matrices of relevant sizes. The number of non-random components is determined by comparing the eigenvalue vector of the empirical data matrix with the vector of mean eigenvalues from the simulations. Only those leading empirical components with eigenvalues greater than their simulated equivalents are retained.

Accordingly, simulations of normal random deviates are performed here. These approximate matrices of Rasch standardized residuals for situations in which the data fit the model. O'Connor's (2000) SAS program was used to efficiently perform multiple simulations.

In Table 1, the average eigenvalues, along with their 5th and 95th percentile values, are presented, obtained from the simulation of different numbers of subjects (N) and items (L). The simulated data are all random noise. The graphs shows the Cattell scree plot for the eigenvalues of the first 20 components.

It is seen that the value of 1.40 is always exceeded by the first eigenvalue, and usually by the second. Consequently, the recommendation is to decide the criterion eigenvalue directly from relevant simulations.

Gilles Raîche
Université du Québec à Montréal
Département d'éducation et pédagogie

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 629-637.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Linacre, J. M. (1998). Detecting multidimensionality: which residual data-type works best? *Journal of Outcome Measurement, 2, 3,* 266-283.

Humphreys, L. G. and Montanelli, R. G. (1975). An examination of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10*, 193-206.

O'Connor BP (2000) SPS, SAS, and MATLAB programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers, 32*, 396-402.

Smith, R. M and Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. *In* M. Wilson (Ed.): *Objective Measurement: Theory into Practice. Volume 2*. Greenwich: Ablex.

Wright, B.D. (1996) Local dependency, correlations and principal components. *Rasch Meas Trans, 10, 3,* 509-511.

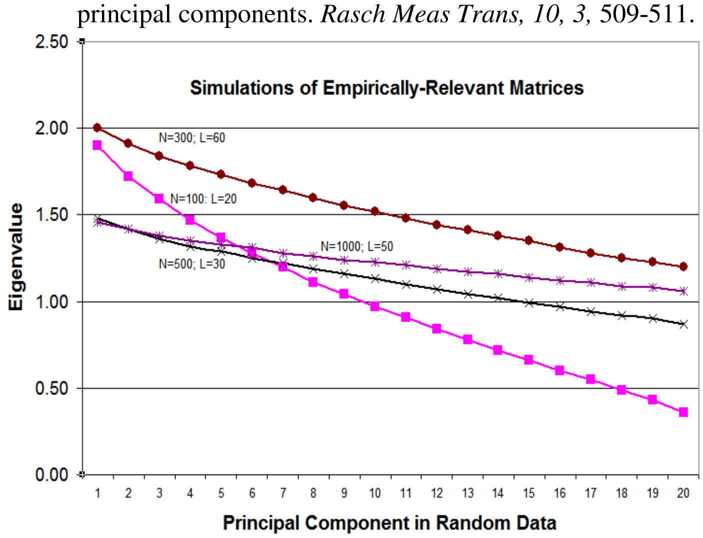


Table 1. Principal component eigenvalues of simulated correlation matrices

	N=100; L=20			N=500; L=30			N=1000; L=50			N=300; L=60		
	5th	Mean	95th	5th	Mean	95th	5th	Mean	95th	5th	Mean	95th
1	1.74	1.90	2.07	1.42	1.48	1.55	1.42	1.46	1.50	1.92	2.00	2.09
2	1.61	1.72	1.85	1.37	1.42	1.47	1.38	1.42	1.45	1.85	1.91	1.98
3	1.50	1.59	1.69	1.33	1.36	1.40	1.35	1.38	1.41	1.79	1.84	1.91
4	1.39	1.47	1.56	1.29	1.32	1.36	1.33	1.35	1.38	1.73	1.78	1.83
5	1.29	1.37	1.46	1.25	1.29	1.32	1.31	1.33	1.35	1.68	1.73	1.78