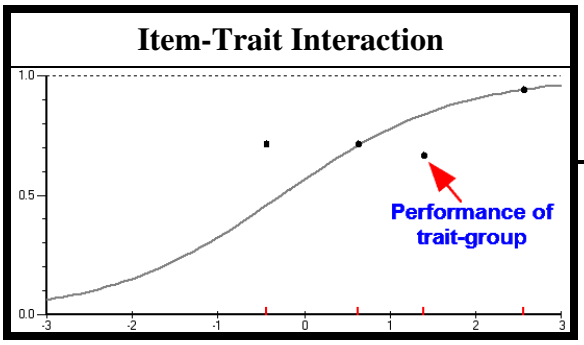


RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

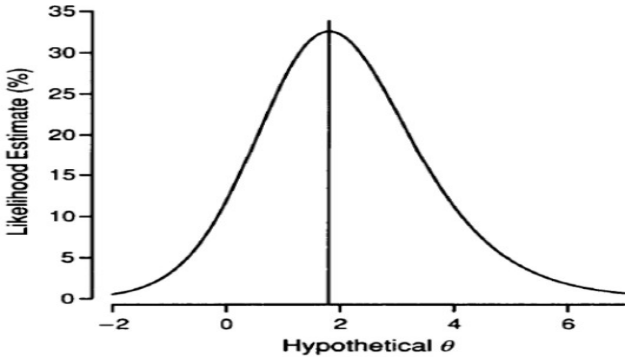
Vol. 21 No. 1 Summer 2007 ISSN 1051-0796

Rasch Measurement and the R Statistics Environment

Rasch measurement has been slow to penetrate into the mainstream of statistical thinking despite positive comments from recognized authority figures such as Otis Dudley Duncan (1984) and Leo Goodman (1990).

But there are promising signs. One is the article "Using the open-source statistical language R to analyze the dichotomous Rasch model" by Y. Li in Behavioral Research Methods, 2006, 38(3), 532-41.

Its Abstract states: "R, an open-source statistical language and data analysis tool, is gaining popularity among psychologists currently teaching statistics. R is especially suitable for teaching advanced topics, such as fitting the dichotomous Rasch model - a topic that involves transforming complicated mathematical formulas into statistical computations. This article describes R's use as a teaching tool and a data analysis software program in the analysis of the Rasch model in item response theory. It also explains the theory behind, as well as an educator's goals for, fitting the Rasch model with joint maximum likelihood estimation. This article also summarizes the R syntax for parameter estimation and the calculation of fit statistics. The results produced by R is compared with the results obtained from MINISTEP and the output of a conditional logit model. The use of R is encouraged because it is free, supported by a network of peer researchers, and covers both basic and advanced topics in statistics frequently used by psychologists."



Li's Figure 2, showing the likelihood function for a 3-item test. This is the only graphical Figure in the article.

Li's article is a competent presentation of Rasch estimation with R. His example dataset is the familiar Knox Cube Test. So his work is a springboard for statisticians looking for a familiar entry point into the somewhat specialized Rasch world.

Li's article also states that "R's pedagogical value makes it well suited for the underlying logic of ... statistical methods. R's design philosophy emphasize data visualization ... These design characteristics not only help students understand the critical abstract theoretical concepts ,,,, they also help students connect abstract statistical concepts with computations."

This is the next step for those following along Li's path. Number crunching is a necessary first step, but picturing the latent variable and conceptualizing what the measures mean are the direction in which the path leads. Let us hope that a subsequent paper on Rasch and R will capitalize on these ideas and include the construction of maps and the underlying measurement concepts which are the motivation behind Rasch models.

Courtesy of William P. Fisher, Jr.

Duncan, O.D. (1984) Rasch measurement: Further examples and discussion. In Charles F. Turner and Elizabeth Martin, editors, Surveying Subjective Phenomena, volume 2, chapter 12, pages 367–403, Russell Sage Foundation, New-York.

Goodman, L.A. (1990) "Total-score models and Rasch-type models for the analysis of a multidimensional contingency table, or a set of multidimensional contingency tables, with specified and/or unspecified order for response categories," Sociological Methodology, Vol. 20, edited by Clifford C. Clogg, Oxford: Basil Blackwell.

Table of Contents

Rasch Measurement With Winsteps

Monday-Tuesday, September 24-25, 2007
Chicago Board Of Education OTS Training Center
15th Floor, 125 S. Clark Street, Chicago, IL

This training emphasizes Rasch measurement concepts and Winsteps rating scale analysis in education with three competency goals:
a) measurement foundations without software,
b) foundations with active software participation, and
c) independent Rasch measurement applications.

Enrollment: For information contact Nikolaus Bezruczko at Nbezruczko@msn.com

Rasch and Continuous Variables

Question: Can Rasch analyze continuous response-level data, such as time and distance?

Answer: There are Rasch models for continuous observations, but processes are rarely truly continuous. Rasch is formulated in terms of distinguishable qualitative advances. How much better, faster, more accurate, does a performance need to be for it to be noticeably better? Think of the same thing in human weight. Our weight varies all day long, so it is not until it has changed by 2 kilos that we really notice a difference. The basic approach in Rasch is to start by categorizing really big increments. If those analyze successfully, we can then reduce the size of the increments until we reach the level where further reduction introduces more randomness than information into the data.

There are several indicators of over-categorization. One is that the model polytomous category probability curves start to look a mess, instead of an advancing range of hills. Another is that, as the number of categories increase, the sample-person "test" reliability falls far behind the value predicted by the Spearman-Brown Prophecy Formula. Going from 2 categories (one decision per item) to 3 categories (two decisions per item) is somewhat like doubling the test length, but not so efficient. So given the reliability, R(m), for an m-category rating scale, we would predict the reliability for an (m+1) category rating scale to be appreciably better, in the range:

$$R(m) < R(m+1) < m*R(m) / (m-1+R(m))$$

We can keep track of the reliability as we increase the number of categories. When reliability not longer shows a reasonable increase (or starts to decrease) we have over-categorized. *John M. Linacre*

Rasch Measurement Workshop

Saturday-Sunday, October 20-21, 2007

University of Illinois at Chicago

This workshop provides a comprehensive introduction to Rasch measurement, dichotomous and polytomous Rasch models, and many-facet Rasch measurement. Workshop directors are Richard Smith and Everett Smith. Workshop attendees will receive a copy of two JAM Press book, *Introduction to Rasch Measurement,* and *Rasch Measurement: Advanced and Specialized Applications,* and a one year subscription to the *Journal of Applied Measurement.* The UIC Campus is a short cab ride from Michigan Avenue hotels.

http://www.jampress.org/

Rasch-related Coming Events

Sept. 2007 - Dec 2008 3-day Rasch courses, (A. Tennant, RUMM) , Leeds, UK, http://home.btconnect.com/Psylab_at_Leeds/

Sept. 24-25, 2007, Mon.-Tues. Rasch Measurement with Winsteps, (N. Bezruczko), Chicago, http://www.rasch.org/bezruczko.htm

Sept. 26-28, 2007, Wed.-Fri. 2a Reunión Regional Norte, Centroamérica y Caribe de Evaluación Educativa, Mexico, http://www.ieesa-kalt.com/forocampeche/foro.html

Oct. 20-21, 2007, Sat.-Sun. Introduction to Rasch Measurement Course, Chicago, (E. Smith & R. Smith, Winsteps, Facets), Chicago, http://www.jampress.org/

Nov. 9 - Dec. 7, 2007, Fri.-Fri. Many-Facet Rasch Measurement online course, (M. Linacre, Facets), www.statistics.com/courses/facets

Jan. 7-11, 2008, Mon.-Fri. Introductory course on Rasch measurement, (Andrich, RUMM), Australia, http://www.education.uwa.edu.au/httpwww.education.uwa.edu.aunews/rasch_conference

Jan. 14-18, 2008, Mon.-Fri. Advanced course on Rasch measurement, (Andrich, RUMM), Australia, http://www.education.uwa.edu.au/httpwww.education.uwa.edu.aunews/rasch_conference

Jan. 21, 2008, Mon. One-day RUMM Workshop, (Andrich, RUMM), Australia, http://www.education.uwa.edu.au/httpwww.education.uwa.edu.aunews/rasch_conference

Jan. 22-24, 2008, Tues.-Thurs. 3rd International Conference on Measurement in Health, Education, Psychology and Marketing: Developments with Rasch models, Australia, http://www.education.uwa.edu.au/httpwww.education.uwa.edu.aunews/rasch_conference

Feb. 15 - March 15, 2008, Fri.-Fri. Practical Rasch Measurement with Winsteps online course, (M. Linacre), www.statistics.com/courses/rasch

March 22-23, 2008, Sat.-Sun. IOMW 2008, New York, http://www.jampress.org/

March 24-28, 2008, Mon.-Fri. AERA Annual Meeting, New York, www.aera.net

May 2-30, 2008, Fri.-Fri. Many-Facet Rasch Measurement online course, (M. Linacre, Facets), ,www.statistics.com/courses/facets

2006-2007 Annual Report for the Rasch SIG

Edward W. Wolfe, Secretary/Treasurer

The Rasch SIG annual business meeting was held during the Annual Meeting of AERA in Chicago, IL on April 10th, 2007. SIG President Tom O'Neill facilitated the meeting. Secretary/Treasurer Ed Wolfe presented the following financial and membership statement:

	Date	Balance
Opening	(April 2006)	$5033.15
Closing	(March 2007)	$6880.15

Expenses included the AERA SIG administration fee ($225) and payment of the website hosting fees ($198). Deposits were made for 138 memberships paid during 2006-2007. Current membership of the SIG is 172.

Upcoming SIG business includes completion of the Program Chair form by 4/13/07. This task has been completed, with Sharon Solloway and Ed Wolfe agreeing to serve as Co-Chairs for the 2008 program. Sharon will serve one more year (served that role in 2007 and will serve in 2008), and Ed will serve for two years (2008 & 2009 programs), with Dimiter Dimitrov agreeing to begin serving a two-year appointment as Co-Chair during the second year of Ed's appointment (for the 2009 and 2010 programs). In addition to this item of business, by 5/11/07 the officers will submit the Officer Designation form to AERA. Tom O'Neill and Ed Wolfe will complete the second year of their two-year terms in 2008.

William Fisher and Sharon Solloway (2007 Program Co-Chairs) reported that they received 37 paper proposals and 2 session proposals. All but 6 of the proposal were reviewed by 3 reviewers. 34 papers were slotted into 6 paper sessions and 2 paper discussion sessions.

Two new items were presented at the meeting. This fall, the SIG will hold elections, so if you'd like to volunteer to serve as the SIG Chair or Secretary/Treasurer, please contact Ed Wolfe (edwolfe@vt.edu) or Tom O'Neill (toneill@ncsbn.org).

Next year IOMW will be held in New York, hosted by Data Recognition Corporation: http://www.jampress.org/

The meeting concluded with a presentation by George Engelhard that described *Historical Trends in Measurement Research.* Following the meeting, the group met at the Manaco Hotel for a reception and dinner hosted by *Pearson VUE.*

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

SIG Chair: Thomas O'Neill, Secretary: Ed Wolfe
Program Chairs: Sharon Solloway & Ed Wolfe
SIG website: www.raschsig.org

Third International Rasch Measurement Conference

Perth, Western Australia

22 - 24 January 2008

Pre-Sessions: Jan. 7-11, 14-18, 21, 2008

Submissions close: Sept. 14, 2007

Topics for the conference:

- Cumulative models for attitude and trait measurement-dichotomous and ordered category models.
- Unfolding models for preference and choice -folding the Rasch models
- Rasch model applications in education (e.g., large scale test equating, benchmarking), psychology (e.g., intelligence testing, linking quantitative and stage developmental data)
- Item banking
- Computer adaptive testing
- Marketing (e.g., pairwise designs for preference and choice studies)
- Health care outcomes (e.g., linking performance scales)
- Using simulation studies for clarifying methodological issues (e.g., tests of fit, measurement precision, effects of multidimensionality and response dependence)
- Developments in Rasch modeling (e.g. differential item functioning)
- Understanding response processes compatible with the Rasch models
- Epistemology, fundamental measurement and Rasch models
- History and philosophy of measurement and Rasch models

January 7-11 Introductory course on Rasch measurement. Includes use of the program RUMM

January 12 Course barbecue

January 14-18 Advanced course in Rasch measurement. Includes use of the programs RUMM, RATEFOLD

January 21 One day workshop focusing on using RUMM

January 22 - 24 Conference papers on applications of Rasch and related measurement models in any substantive field of application - education, psychology, health care and rehabilitation, marketing, etc.

January 22 *Conference dinner* at the Nedlands Golf Club, located two miles from the city of Perth, and overlooking the Swan River.

http://www.education.uwa.edu.au/httpwww.education.uwa.edu.aunews/rasch_conference

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

Campeche, Mexico, 26-28 September 2007

sponsored by the Institute of Evaluation and Advanced Engineering

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

http://www.ieesa-kalt.com/forocampeche/foro.html

Español Spanish Language Meeting: PROGRAMA

Hora	*Día 26 de Septiembre de 2007* **ACTIVIDAD**	
14:00 - 15:30	REGISTRO Y ENTREGA DE MATERIALES Responsables: Coordinadores Lugar: Lobby del Centro de Convenciones	
16:00 - 18:30	TALLERES 1. Introducción al diseño de pruebas objetivas. Responsable: Agustín Tristán. Salón 5 2. Diseño de tablas de validez de contenido. Responsable: Gerardo Gutiérrez. Salón 6 3. Taxonomías educativas. Responsable: Deyanira Molgado. Salón Ejecutivo	
20:00 - 22:00	CENA DE BIENVENIDA Lugar: Ex templo de San José	
Hora	*Día 27 de Septiembre de 2007* **ACTIVIDAD**	
8:00 - 9:00	REGISTRO Y ENTREGA DE MATERIALES Responsables: Coordinadores Lugar: Lobby del Centro de Convenciones	
9:00 - 10:00	INAUGURACION Responsables: Agustín Tristán/ IEIA Norma Lozano/ Colegio de Bachilleres Campeche Autoridades (declaratoria inaugural) Lugar: Salón 2 y 3 Centro de Convenciones	
10:00 - 11:30	CONFERENCIA MAGNA "Enfoques Alternativos en la Evaluación de los Aprendizajes. Hacia un Modelo de Gestión de la Evaluación" Responsable: Dr. Joan Mateo Andrés/ Universidad de Barcelona Lugar: Salón 2 y 3 Centro de Convenciones	
11:30 - 14:00	MESAS DE TRABAJO	
	Educación Media Superior	Proyectos Especiales
	1.1 Diseño y Aplicación del Programa Complementario de Evaluación Educativa en Educación Media Superior en Sonora Ponente: Amado A. Montoya/ IEEES Lugar: Salón 5 1.2 Comparación de Estilos de Evaluación del Aprendizaje entre el Bachillerato General y la Educación Superior Ponente: Javier Tarango/ COBACH Chihuahua Lugar: Salón 6	2.1 Un Sistema Integral de Evaluación de la Educación en el Distrito Capital (Bogotá) Ponente: Edilberto Novoa/ SED Colombia Lugar: Salón 6 2.2 Evaluación del Programa Escuelas de Calidad en el Estado de Puebla Ponentes: Jacqueline Herrera/ SEP y León R. Garduño/ Universidad de las Américas de Puebla Lugar: Salón 6
14:00 - 15:30	C O M I D A	
15:30 - 18:00	MESAS DE TRABAJO	
	Educación Superior	Educación Primaria
	3.1 Hacia un Modelo de Evaluación y Certificación del Ejercicio Profesional: el Caso de Medicina Veterinaria y Zootécnica Ponentes: Rafael Hernández y Sonia Rivera/	4.1 Evaluación de la Comprensión en Ciencias Naturales de niños y niñas de los grados 5º y 9º Ponente: Álvaro García, Jairo R. Pinilla y Olga R. Rodríguez/ Colombia

	CENEVAL Lugar: Salón 5 3.2 Proceso de Admisión Certificado Ponente: Ma. Del Carmen Hernández, Fernando Cuevas y Francisco Javier Martínez/ UASLP Lugar: Salón 5		Lugar: Salón 6 4.2 Evaluación del Diagnóstico para determinar el nivel de Aprendizaje de Lengua Escrita y las Matemáticas. 1º y 2º Ponente: Rodrigo Álvarez/ USEBEQ Lugar: Salón 6
20:00	ESPECTÁCULO DE LUZ Y SONIDO DE LA PUERTA DE TIERRA		
Hora	*Día 28 de Septiembre de 2007* **ACTIVIDAD**		
9:00 - 10:30	CONFERENCIA MAGNA "Plan de Trabajo del Segundo Estudio Regional Comparativo y Explicativo, SERCE" Responsables: Dr. Daniel Bogoya, Lilia Toranzos, Giuliana Espinosa y Mauricio Castro, Colombia Lugar: Salón 2 y 3		
10:30 - 14:00	MESAS DE TRABAJO		
	Nuevas Tendencias	Evaluación Docente	Análisis Psicométricos
	5.1 El Certificado Internacional de Español. Un Proyecto de la UNAM y el Instituto Cervantes de España para certificar el dominio de la lengua española en el mundo. Ponente: Martha Jurado y Claudia Cárdenas/ UNAM Lugar: Salón 5 5.2 Examen en Línea Ponente: Miguel Ángel Ylizaliturri / IEIA Lugar: Salón 5 5.3 Exámenes Departamentales en la UIA. Reglamentación, Diseño, Aplicación en Línea y uso de Resultados Ponentes: Alberto Segrera, Elsa Sánchez, Antonio Miranda y Ma. De Lourdes Caudillo/ UIA Lugar: Salón 5	6.1 Diseño de un Instrumento para Evaluar el Conocimiento de Docentes Acerca de la Enseñanza Constructivista Ponente: Edith Cisneros/ UAY Lugar: Salón 6 6.2 Dictamen de Perfiles Docentes a Partir de la Opinión de los Alumnos. Ponente: Deyanira Molgado/ IEIA Lugar: Salón 6 6.2.1 La experiencia Salvadoreña en Evaluación Docente Ponente: Ana Julia Martínez/ IEIA El Salvador Lugar: Salón 6 6.3 Sistemas de Evaluación de Profesores: caso del Colegio de Bachilleres del Estado de Campeche Ponente: Matilde Salazar/ COBACH Campeche Lugar: Salón 6	7.1 Desarrollo de un Modelo de Eficacia Educativa: Estudio Multinivel en un Subsistema de Educación Superior Ponentes: Sonia Rivera y Rafael Hernández/ CENEVAL Lugar: Salón Ejecutivo 7.2 Consideraciones Psicométricas sobre el papel del Contexto en Preguntas de Conocimientos: Una Aproximación Empírica Ponente: Lady C. Lancheros/ Universidad Nacional de Colombia Lugar: Salón Ejecutivo 7.3 Equiparación de Puntuaciones con TRI y TCT en una Prueba de Ingeniería Ponente: Olga Rodríguez/ Universidad Nacional de Colombia Lugar: Salón Ejecutivo
14:00 - 15:30	C O M I D A		
15:30 - 17:00	MESA DE TRABAJO		
	"Calificación de Pruebas con Reactivos de Respuesta Abierta" Lugar: Salón 5 Responsable: Martha Rocha y Carlos Pardo/ ICFES		
17:00 - 18:00	"Software, Análisis y Calificación de Textos en Español" Lugar: Salón 2 y 3 Responsable: Agustín Tristán López		
18:00 - 18:30	CLAUSURA Lugar: Salón 2 y 3 Responsables: Agustín Tristán López (IEIA), Norma Lozano Reyes (COBACH), Autoridades (palabras de clausura)		

Living Capital Metrics

Irving Fisher is one of several economists who are together largely responsible for our common assumption today that markets consist of the aggregate of all of our individual private trades and exchanges, which, taken altogether, set prices (Boumans, 2001). Rasch's separability theorem may be indebted in some as yet unknown way to I. Fisher's separation theorem, since the latter was a close colleague of Ragnar Frisch, one of Rasch's teachers and a winner of the Nobel Prize in Economics. In addition, both I. Fisher and Frisch were involved in the founding of the Cowles Commission, where Rasch was a scholar in residence in 1947, in Chicago. At the very least, one might expect that the force of Frisch's reaction to Rasch's "disappearing parameter" (Andrich, 1997; Wright, 1980) might have been less vigorous if he had not already been exposed to the idea in I. Fisher's work.

One interpretation of the widespread successful application of Rasch's models could be that the market principle does in fact function in the economies of a wide variety of other forms of capital, such as human, social, and natural (W. Fisher, 2002), contrary to the esteemed opinions of others unfamiliar with these applications (Arrow, 1963). As the long-term profitability of socially responsible investing and environmentally sustainable management practices becomes increasingly apparent (see recent issue of Business Week, the Economist, and others for more information), so also will the need to expand local economies of scale-dependent metrics to regional and global economies of invariant reference-standard metrics.

As De Soto (2000) shows, the mystery as to why capitalism works in some countries and not others has a great deal to do with the existence of legal and financial infrastructures that universally recognize and accept certain instruments, currencies, and other forms of transferable representations (titles, deeds, etc.) as valid conveyances of value. Trillions of US dollars of value lie unusable and dead globally within various national economic systems lacking the appropriate infrastructure, though this is changing as the World Bank, governments, and non-governmental organizations mount programs for building the needed institutions and processes.

Manufactured and liquid capital, and property, have successfully been brought to life in Western economies, but human, social, and natural capital remain dead, or as yet unborn, tied as they are virtually everywhere to nontransferable representations—scales with values that change depending on local particulars.

Because of the inexorable force of economic globalization, the day is inevitably approaching when measures built from Rasch models will be incorporated in the definitions of every kind of fungible human, social, and natural capital metric. In much the same way that price and value information have been used for centuries, these living capital metrics will be used by consumers to make purchasing decisions, by investors to make financing decisions, by executives to make resource allocation decisions, by managers to make quality improvement decisions, and by accountants to make earnings and profit statements.

To appreciate the scale of these applications, consider the fact that some estimates of today's currently existing human, social, and natural capital resources put their value at 99% or more of the global economy. Though there may be something humorous about assigning monetary value to the air and water purification services essential to life, the value of the ongoing services provided by natural capital alone every year is estimated to be about equal to the annual gross world product (Hawken, Lovins, & Lovins, 1999, p. 5). If the value produced by natural capital were interest paid on invested assets at the rate of 10% annually, the world's natural capital resource stock would then have a value ten times the value of the annual gross world product.

Similarly, the World Bank estimates the sum value of global human capital to be three times more than the existing capital values included in standard accounting balance sheets (Hawken, Lovins, & Lovins, 1999, p. 5). The bottom line is that the current state of capitalism is so incomplete that its accounting methods are dealing with only a tiny fraction of the actual value of the available resources, and almost all of that is in the form of manufactured and financial capital and property.

And all of this comes to bear as many sectors of the economy are struggling to find new untapped sources of inefficiency that could be mined for profits. Given the seemingly endless inflationary spirals in the economies of education, health care, and social services, being able to grow living human, social, and natural capital in socially responsible ways will likely open up huge new markets with potentials defined less by short-term profiteering than by long-term sustainability.

When this happens, research and technologies that respond to the demand for what Irving Fisher called numerical indexes conforming with his separation theorem, or what Rasch called measures conforming with his separability theorem, will be in the mainstream of research in the human sciences, instead of at the periphery, where they are today. The shift from today's scales defined from within the positivist statistical paradigm of descriptive models to those defined from within a post-positivist measurement paradigm of prescriptive models will reach its tipping point when investors, accountants, managers, and consumers all take for granted metrics capable of functioning as common currencies for the exchange of human, social, and natural capital.

If history is any guide (Latour, 2005; Rabkin, 1992), this shift will not take place as a result of academic exercises

in theory or research. Instead, it will take place only as a result of the accumulated production of actual value, when the repeated utility of separable parameters in the measurement of living capital really facilitates improved quality of life and enhanced accountability. And when these kinds of desired values are reliably reproducible, then, and only then, will there be a decisive shift in the laboratory values that are incorporated into research designs (Daston, 1992; Hunt, 1994; Schaffer, 1992) and in the review criteria applied to publications and funding awards.

The key factor lies in making any given unit of measurement seem natural, as a property of the thing itself, instead of as an artifact of a particular methodology, person, organization, or nationality. And nothing, absolutely nothing, can exert as much power in this regard as a standards laboratory (Latour, 1987, pp. 247-57; Schaffer, 1992). Metrologically equating all brands or configurations of instruments that actually measure the same thing so that they do so in the same unit has historically been the means by which our conception and perception of the naturalness of nature has been socially constructed.

We now have the means for reproducing in the social sciences the successes of the natural sciences in this regard (Fisher, 2000). History can provide another lesson concerning the consequences of efficient capital measurement. Europe rose to global power between 1250 and 1600 by unifying mathematics and measurement in a quantitative model of the world. Because of this model, Europeans "were able to organize large collections of people and capital and to exploit physical reality for useful knowledge and for power more efficiently than any other people of the time" (Crosby, 1997, p. x). In the coming age, the dominant power in the world will be the one that learns to organize human, social, and natural capital more effectively and efficiently than others. Whether this will be done in a manner that respects human rights and democratic principles remains to be seen, but it will be done, in any case.

William P. Fisher, Jr.

Andrich, D. (1997). Georg Rasch in his own words [excerpt from a 1979 interview]. Rasch Measurement Transactions, 11(1), 542-3.
http://www.rasch.org/rmt/rmt111d.htm

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. American Economic Review, 53, 941-73.

Boumans, M. (2001). Fisher's instrumental approach to index numbers. In M. S. Morgan & J. Klein (Eds.), The age of economic measurement (pp. 313-44). Durham, North Carolina: Duke University Press.

Crosby, A. W. (1997). The measure of reality: Quantification and Western society, 1250-1600. Cambridge: Cambridge University Press.

Daston, L. (1992). Baconian facts, academic civility, and the prehistory of objectivity. Annals of Scholarship, 8, 337-363. (Rpt. in L. Daston, (Ed.). (1994). Rethinking objectivity (pp. 37-64). Durham, North Carolina: Duke University Press.)

De Soto, H. (2000). The mystery of capital: Why capitalism triumphs in the West and fails everywhere else. New York: Basic Books.

Fisher, W. P., Jr. (2000). Objectivity in psychosocial measurement: What, why, how. Journal of Outcome Measurement, 4(2), 527-563.

Fisher, W. P., Jr. (2002, Spring). "The Mystery of Capital" and the human sciences. Rasch Measurement Transactions, 15(4), 854
http://www.rasch.org/rmt/rmt154j.htm

Hawken, P., Lovins, A., & Lovins, H. L. (1999). Natural capitalism: Creating the next industrial revolution. New York: Little, Brown, and Co.

Hunt, B. J. (1994). The ohm is where the art is: British telegraph engineers and the development of electrical standards. In A. van Helden, & T. L. Hankins (Eds.), Instruments [Special issue]. Osiris: A Research Journal Devoted to the History of Science and Its Cultural Influences, 9, 48-63. Chicago, Illinois: University of Chicago Press.

Latour, B. (1987). Science in action: How to follow scientists and engineers through society. New York: Cambridge University Press.

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Clarendon Lectures in Management Studies. Oxford, England: Oxford University Press.

Rabkin, Y. M. (1992). Rediscovering the instrument: Research, industry, and education. In R. Bud & S. E. Cozzens (Eds.), Invisible connections: Instruments, institutions, and science (pp. 57-82). Bellingham, Washington: SPIE Optical Engineering Press.

Schaffer, S. (1992). Late Victorian metrology and its instrumentation: A manufactory of Ohms. In R. Bud & S. E. Cozzens (Eds.), Invisible connections: Instruments, institutions, and science (pp. 23-56). Bellingham, WA: SPIE Optical Engineering Press.

Wright, B. D. (1980). Foreword, Afterword. In Probabilistic models for some intelligence and attainment tests, by Georg Rasch (pp. ix-xix, 185-199.
http://www.rasch.org/memo63.htm

False Concreteness

Peter Drucker described years ago ... "the illusion of false concreteness." This is believing that a result or recommendation presented with figures showing multiple decimal places is somehow more accurate than a judgment call presented in words alone.
www.mondaymemo.net/011015feature.htm

Warm (Maximum) Likelihood Estimates

Ronald A. Fisher (1922) formulated the concept of the likelihood of the data given a statistical model that is hypothesized to have generated those data and a set of parameter estimates. Maximum likelihood estimates are the parameters values which maximize the likelihood that the observed data would have been generated. Thus MLE values correspond to the mode of the likelihood function.

Thomas Warm (1989) points out the that those modal estimates are biased when viewed from the likelihood function as whole. He suggests that, rather than the mode of the likelihood function, estimates should be based on its mean. These estimates have come to be called "Warm estimates", and his approach is Warm (or Weighted) (Maximum) Likelihood Estimation" (WLE or WMLE).

In the Rasch model, the estimation of MLE and WLE require iteration. WLE is more computationally intensive. Warm demonstrates that the asymptotic variance of MLE and WLE estimates are the same, meaning that the estimates have the same model standard errors.

WLE estimates are generally slightly more central than MLE estimates, though the implications of this for practical applications are not clear, because the difference is usually less than the standard error of the estimates.

Fisher R.A., (1922) "On the mathematical foundations of theoretical statistics" Philosophical Transactions of the Royal Society of London (A) 222, 1922, p. 309-368.

Warm T.A. (1989) "Weighted Likelihood Estimation of Ability in Item Response Theory." Psychometrika, 54, 427-450.

Raw Score-to-Measure (Scaled Score) Tables

Question: Is it good practice for me to supply a lookup table or Excel sheet to convert sum scores into the measure scores so that users get their results on an interval scale? Is there a good example to cite?

Answer: Many publishers and authors supply a "raw score to scale score" conversion chart. It is convenient for the users, and also allows different versions and revisions of the instrument to continue to generate comparable scale scores. Here is an example showing raw scores and their Rasch logit equivalents:

Raw Score	Rasch-Transformed Score	SE
0	-9.46	1.52
1	-8.56	1.18
2	-7.40	1.01
3	6.38	1.04

Hsueh I-P., Wang W.-C., Sheu C.-F., and Hsieh C.-L. (2004) Rasch Analysis of Combining Two Indices to Assess Comprehensive ADL Function in Stroke Patients. Stroke, 35(3): 721 - 726.

But it is more useful to publish a transformation of Rasch logits, not the logits themselves. Transform the logits such that all Rasch "scores" are reported as positive integers. Here is a rescaled example. You can note the exact logit transformation as a footnote to the Table if it will be meaningful to your readers.

RAW SCORE TO SCALE SCORE CONVERSION TABLE
MARCH 2002 GRADE 4

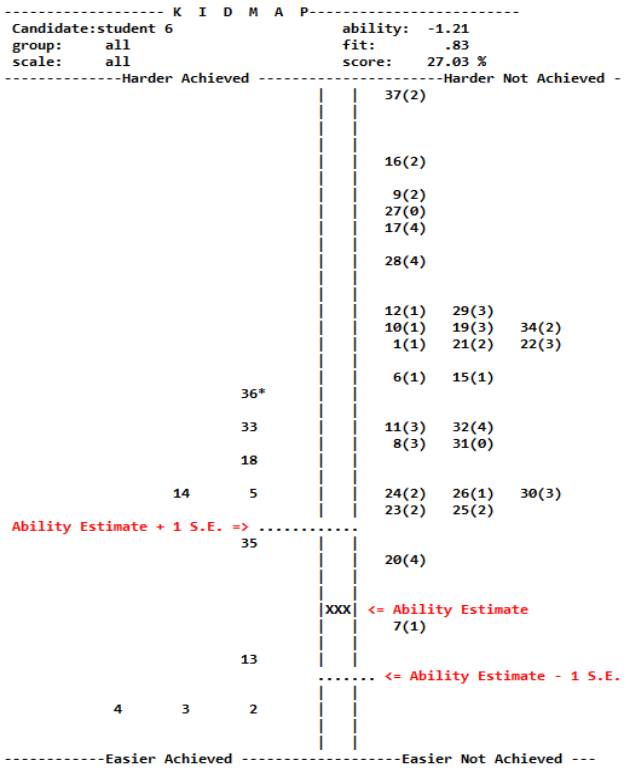
Reading		Mathematics		Citizenship	
Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score
1	109	1	66	1	88
2	122	2	87	2	106
3	130	3	100	3	117
4	136	4	110	4	125
5	140	5	118	5	131
6	144	6	125	6	136
7	147	7	131	7	141
8	151	8	137	8	145
9	153	9	142	9	149
10	156	10	147	10	152

Ohio Department of Education, Office of Assessment

KIDMAPs via Excel

Tsair-Wei Chien of the Chi-Mei Medical Center, Taiwan has made available an Excel add-in which constructs KIDMAPs for individual examinees. These Excel routines for dichotomous and polytomous data call Winsteps to compute persons measures, etc. The Excel routines are designed for use by classroom teachers.

www.healthup.org.tw/leapfrog/kidmap_description.htm



Clinical Significance

"Evaluations of the outcomes of psychological treatments are favorably enhanced when the published report includes not only statistical significance and the required effect size but also a consideration of clinical significance."

Kendall, P.C. (1997). Editorial. *Journal of Consulting and Clinical Psychology, 65, 3-5.*

Rating Scale Instrument Quality Criteria					
Criterion	**Poor**	**Fair**	**Good**	**Very Good**	**Excellent**
Targeting	> 2 errors	1-2 errors	< 1 error	< .5 error	< .25 error
Item Model Fit Mean-Square Range Extremes	< .33 - >3.0	.34 - 2.9	.5 - 2.0	.71 – 1.4	.77 – 1.3
Person and Item Measurement Reliability	<.67	.67-.80	.81-.90	.91-.94	>.94
Person and Item Strata Separated	2 or less	2-3	3-4	4-5	>5
Ceiling effect: % maximum extreme scores	>5%	2-5%	1-2%	.5-1%	<.5%
Floor effect: % minimum extreme scores	>5%	2-5%	1-2%	.5-1%	<.5%
Variance in data explained by measures	<50%	50-60%	60-70%	70-80%	>80%
Unexplained variance in contrasts 1-5 of PCA of residuals	>15%	10-15%	5-10%	3-5%	<3%

This Table has been developed by William P. Fisher, Jr. based on the Rasch literature and his many years of experience conducting Rasch analyses in different settings.

RUMM2020 Item-Trait Chi-Square and Winsteps DIF Size

Obtaining equivalent numerical results from different software packages can be challenging. Item-trait interactions are an example. The RUMM2020 Item Fit Table shows the item-fit output for item I0104 from a RUMM2020 analysis. The *Location* is the Rasch item difficulty estimate with *SE* being its standard-error precision. The *FitResid* is the standardized sum of squared residuals with *DF* being its estimated degrees of freedom. *FitResid* is equivalent to the standardized OUTFIT statistic of Winsteps.

The *ChiSq* is the item-trait interaction. In this example the latent trait is stratified into four class intervals each containing a trait-group of approximately one quarter of the total person sample. Since there are 4 intervals, there are three degrees of freedom, *DF*, for the chi-square as indicated. The chi-square is computed from a comparison of the observed overall performance of each trait-group on the item with its expected performance. This quantifies the size of the departure of the empirical item characteristic curve from its model values, so identifying the magnitude of the item-by-trait (item-by-ability level) interaction for this item. *Prob* reports the statistical probability of observing the chi-square value (or worse) when the data fit the Rasch model. In this example, the chi-square has 3 degrees of freedom and so has an expected value of 3.0. Its observed value is 21.707, with a probability of that value of larger being observed by chance of only 0.000076. So we would reject the null hypothesis that the overall performance of the trait-groups fits the Rasch model. We are observing an item-trait interaction for item I0104.

This item-trait chi-square is featured in RUMM2020 documentation as an indicator of item behavior, more so than the FitResid, but there is no obviously equivalent statistic currently reported by Winsteps. This can be awkward when research teams are employing both software packages. Here is how to generate the equivalent statistic in Winsteps:

1. Decide on the number of trait-groups. 4 here.

2. Order the persons by measure (location). Writing the person-measure PFILE to Excel facilitates these steps.

3. Omit extreme scores. These cannot show an interaction.

4. Stratify the person-ability range into trait-groups of as equal numerical size as possible, keeping all persons with the same measure in the same group.

4. Number the trait-groups and put the trait-group number into each person label.

5. Perform a DIF analysis of item by trait-group-number.

6. Obtain the *t*-statistic for each item-trait DIF effect.

7. For each item, square and sum the *t*-statistics for the item-trait groups. This is the RUMM2020 chi-square.

8. The chi-square d.f. is the count of trait-groups less one.

In our example, the Winsteps DIF Table shows each trait-group as a *Person Class.* The *Observations Count* is the number of persons in the group. *Average* is their average rating. *Baseline Expect* is the expected value of the *Observations Average*. *Measure* is the item difficulty measure corresponding to the *Baseline Expect* rating on this item, Item 104. It is expected to be the same for every trait-group. The *DIF Score* is the difference between the *Observations Average* and *Baseline Expect ratings*. The *DIF Measure* is the item difficulty that would produce the *Observations Average*. So that *DIF Size* is the difference between the *Baseline Measure* item difficulty and the item difficulty observed for this group, the *DIF Measure. S.E.* is the standard error of the *DIF Size*. The *t*-statistic is a hypothesis test that the *DIF Size* is due to chance alone, it is the *DIF Size* divided by its *S.E.*

The Winsteps *t*-statistic is approximately a unit-normal deviate. Squaring and summing the four of these for item I0104 amounts to 20.05, close to the RUMM2020 ChiSq of 21.707. Thus this procedure yields approximately the same number as the RUMM2020 ChiSq. Over 72% of the Winsteps chi-square is contributed by the 4th trait-group, indicating that the item-trait interaction is primarily due to the unexpectedly poor performance by the high ability group.

These statistics are sensitive to the number of item-trait groups, so verify that an item is defective (from an item-trait perspective) by replicating this process with different numbers of item-trait groups.

John M. Linacre

RUMM2020 Item Fit Table

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob
104	I0104	Poly	0.246	0.137	2.852	228.56	**21.707**	3	0.000076

Winsteps DIF Table

Person Class	Observations		Baseline		DIF					Item	
	Count	Average	Expect	Measure	Score	Measure	Size	S.E.	t	Number	Name
1	57	0.53	0.40	0.25	0.12	-0.27	-0.52	0.27	**-1.92**	104	I0104
2	55	0.62	0.55	0.25	0.07	-0.04	-0.29	0.28	**-1.06**	104	I0104
3	59	0.68	0.62	0.25	0.05	0.02	-0.24	0.28	**-0.85**	104	I0104
4	60	0.47	0.70	0.25	-0.23	1.24	0.99	0.26	**3.81**	104	I0104