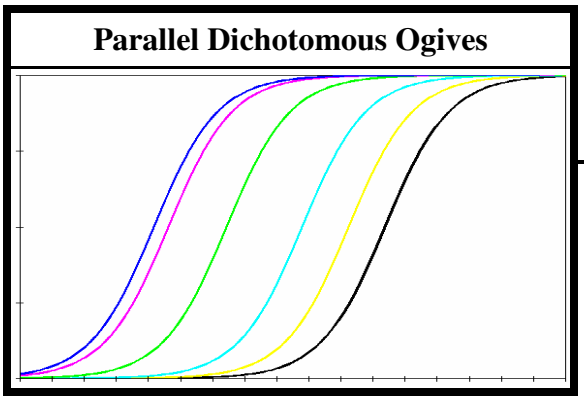

RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 18 No. 4 Spring 2005 ISSN 1051-0796

Taking A Byte Out of Job Satisfaction

Computerized surveys have all but replaced traditional paper and pencil instruments in many organizations (Good, 1997). The economic benefits associated with reduced printing and postage costs, the speed of data collection and the wide availability of easy to use survey programs has fueled the surge in computer based survey popularity. A wealth of research has demonstrated the comparability of computer and written examination formats across achievement and aptitude testing, but few researchers have compared the results of *satisfaction* measured gathered using the different techniques (Comley, 1998).

The present study compares the results of satisfaction surveys administered in a large, public organization. This study replicated my earlier investigation in a similar governmental organization. The sample included 832 employees who agreed to participate in the survey process. Two identical survey instruments were created and delivered to each employee over two consecutive weeks. In earlier research, the individuals completing the written survey were different than those completing the computerized survey. I our study, each respondent completed each version of the instrument. To overcome possible completion order difficulties, half completed the computer form first, while the other half completed the written form first. It was also hoped that this approach and the sizable number of participants involved would overcome the real changes in employee satisfaction that may occur from one week to the next.

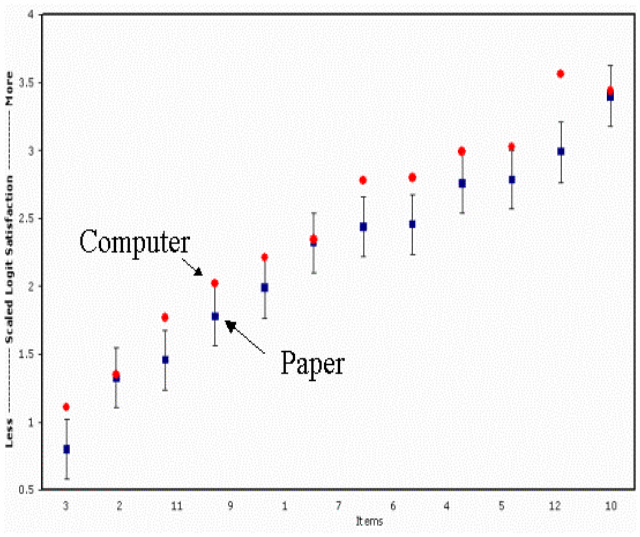

Table 1: Items Manifesting Significant Differences: Computer vs. Paper Delivery

3) I am satisfied with the benefits I receive
4) Teamwork is encouraged
6) My supervisor allows me to contribute in managerial decision making
10) My immediate supervisor is friendly and helpful
11) I feel I have job security

A uniquely designed set of 12 satisfaction items were created for this experiment. The items covered a variety of satisfaction related factors, including compensation, supervisory and collegial relations, environment, etc.

Data from the written survey were analyzed first and baseline logit difficulties defined. Next data from the computer administered surveys were anchored to written item difficulties (items 2,7) to generate a simple, comparable set of computerized delivery item difficulties.

Results were quite striking. Five of the ten non-anchored items were found to be significantly different based on instrument delivery format. Furthermore, there was a clear trend across the instrument. Overall, respondents tended to rate themselves as more satisfied when responding to the computerized version.

To better understand the observed differences in

Table of Contents

satisfaction, three focus groups were convened. Attendees included 30 employees who volunteered to discuss the experience in a confidential forum. The single common response across members and groups appeared to be comfort with the level of confidentiality. Respondents felt that their answers would not be traceable to them in paper format, but were not convinced of the same security using the computer. As one female employee stated, "they tell us they monitor our computer use – you know, to stop us from playing games on the internet – and, well, just because they say our responses are confidential, who knows. I ain't risking my job for this thing."

In our discussions, the focus group employees did not report substantive changes in satisfaction, and while this cannot exclude the possibility that changes occurred, those changes should be mitigated by the method employed.

While the result of this single evaluation appears to suggest there may be skewed responses to computerized surveys, it may simply be unique to this particular organization or others like it. On the other hand, the discovery of a difference does emphasize that in a world of efficient survey administration, we cannot take for granted that delivery format is unrelated to outcome. Surveys are not examinations and those conducted in atmospheres with established hierarchies, such as job satisfaction surveys, may carry with them elements of discomfort more demonstrable in a computerized format.

Gregory Ethan Stone
The University of Toledo

Comley, Pete. (1998). On-Line Research: Some Options, Some Problems, Some Case Studies. In Westlake, Andrew et al. (Eds), *New Methods in Survey Research 1998.* Proceedings of the ASC international conference, a satellite meeting for COMPSTAT 98.

Good, K. (1997). A study of factors affecting responses in electronic mail surveys. Dissertation, Western Michigan University, DAI, volume 58-10A, 119 pages.

"It is commonly believed that innovations create changes - but few ever do. Successful innovations exploit changes that have already happened."

Peter Drucker

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

SIG Chair: Randy Schumacker, Secretary: Steve Stemler
Program Chair: Trevor Bond

A Very Rasch prediction - correct so far!

Here is a report that confirms the findings of Boone & Gabel, "A U-turn on the Information Highway?", RMT 8:3, p.369, 1994.

Why Computers Have Not Saved The Classroom

What impact has computer technology had on public education in the US? That's the question journalist Todd Oppenheimer sets out to answer. His conclusion: Putting computers in classrooms has been almost entirely wasteful, and the rush to keep schools up-to-date with the latest technology has been largely pointless, reports Bob Blaisdell. "At this early stage of the personal computer's history, the technology is far too complex and error prone to be smoothly integrated into most classrooms," Oppenheimer writes. "While the technology business is creatively frantic, financially strapped public schools cannot afford to keep up with the innovations." Of course, this is not the first time US schools have been seduced by new technology, Oppenheimer points out. He summarizes the history of technological innovations in American schools and explains how each (TV among them) has been hailed as education's savior. Oppenheimer examines individual schools where technology has been useful, but there he largely credits the enthusiasm and devotion of individual teachers. The most effective teachers, he argues, are those who know enough to ignore the latest technological products and rely on such hands-on technology as pens and paper, musical instruments, wooden blocks, and rulers. These findings contrast sharply with education advocates who argue that education will become increasingly digital, mobile, and virtual.

Excerpted from:
www.csmonitor.coin/2003/1014/p20sO2-lecl.html
Reported in *PEN Weekly NewsBlast, Oct. 17, 2003*

Book Review:
Introduction to Rasch Measurement: Theory, Models, and Application

Published by JAM Press (2004), ISBN: 0-9755351-1-0. www.jampress.org

This comprehensive anthology of 701 pages with contributions from the *Journal of Applied Measurement* and a few other chapters is edited by Everett Smith and Richard Smith.

First, this book fills a clear and important need in the understanding and development of Rasch measurement. Almost all of us have suggested to someone that they get a copy of the 25 year old *Best Test Design* or keep a file drawer of articles on particular topics. Recently, we've had the readable contribution of Bond and Fox; *Applying the Rasch Model,* but it's very timely that someone put the best thought from last 2 decades of work into a primary source. Whether used for university instruction or as a desk reference, I think the book is extremely valuable for our times.

Richard and Everett have divided the text into 3 sections: Theory, Models, and Applications. Being good Rasch practitioners, they have also logically ordered chapters within sections with beginning concepts first so that chapters, even though from different authors, follow each other well. There are subject and author indexes, and all chapters have references. The book is extremely well-presented with clear figures, consistent presentation of equations, and leveled headers. In short, it's pretty easy to navigate the book looking for any particular topic, and I applaud the editors for the organization.

I think this book is just the right thing for those at the beginner stage who want to increase their understanding of Rasch models, or test developers who face an applied problem which they don't know how to handle. Along with the examples in the book and the references to modern software (Winsteps, Facets, RUMM, and Conquest), most test developers will find an answer or explanation to interpretation of all the most common Rasch applications. As a reader for graduate students, the book is seriously under-priced for those students who often pay $50 for something from the copy shop that's poorly printed or $120 for a textbook that is hardly readable!

Because the book consists of different authors writing originally for a journal, there are a few inconsistencies in the use of formula conventions, and some short-changed topics that beg for more discussion, but overall these are minor issues. In fact, the editors have demonstrated their awareness of the field by choosing the best authors for topics that fit (pun intended) the best thinking of those individuals: methods by Linacre, fit statistics by R. Smith, controversies by Andrich, standard setting by Stone.

Novices to the Rasch model won't necessarily notice or know the chapter authors, but they may wonder why popular names in the references aren't represented as chapter authors. By this logic, the book could have benefited with some chapters from some missing authors who didn't have contributions in JAM, but may have been logical additions for a basic collection of this type: Engelhard, Fisher, Masters, etc. Hopefully, the editors will make efforts to get some of those others into the new collection which they are working on now.

Overall, this is an excellent collection of excellent writing. The book is a logical overview of the Rasch model as it stands today that doesn't exist in this detail elsewhere. *Introduction to Rasch Measurement* should be useful for library collections, as a desk reference for test developers, or as a course supplement.

Steve Lang
University of South Florida St. Petersburg

"If an elderly, but distinguished, scientist says that something is possible, he is almost certainly right, but if he says it is impossible, he is very probably wrong."
Arthur C. Clarke, 1969

AERA Annual Meeting: Rasch-related Papers

Montreal, Quebec, Canada

Monday, April 11, 2005

2:15 p.m. - 3:45 p.m. Le Centre Sheraton Montreal / Salon 7

Educational Applications of Rasch Measurement

Chair: Thomas R. O'Neill, National Council of State Boards of Nursing

Bachelor of Education Admissions Procedures: A Many-Faceted Rasch Partial Credit Model Analysis, Peter D. MacMillan, Colin Chasteauneuf

Examining Item Bias on the Third International Mathematics and Science Study Using the Basic Rasch Model, Tenisha Tevis, Regina J. Deil-Amen

Monitoring the Effectiveness of New York's Written Composition Test in English (WCTE) Using Multi-Facet Rasch Measurement, Stephen Hetherman, Madhabi Chatterji

Validation of a Teacher Practice Survey with the Rasch Model, Susan M Gracia

4:05 p.m. - 5:35 p.m. Hilton Montreal Bonaventure / Fontaine, Section A

Contemporary Issues in Psychometric Research

includes:

WINIRT: A Windows-Based Item Response Theory Data Generator With an Equating and DIF Simulation Guide, Hua Fang, George A. Johanson

Aberrant Response Patterns: Issues of Internal Consistency and Concurrent Validity, Iasonas Lambros Lamprianou, Thekla Afantiti-Lamprianou

6:15 p.m. - 7:45 p.m. Fairmont The Queen Elizabeth / Chaudiere

Rasch Measurement SIG Business Meeting

SIG Chair: Randall E. Schumacker.

SIG Program Chair: Trevor Bond

Jean-Guy Blais discusses the impact of Rasch Measurement in Francophone countries.

Tuesday, April 12, 2005

8:15 a.m. - 8:55 a.m. Marriott Montreal Chateau Champlain / Salle de Bal Ballroom & Foyer

Comprehensive School Reform Implementation and Outcomes

includes: Analyzing CSR Implementation with the Rasch Model, Susan M Gracia

10:35 a.m. - 12:05 p.m. Le Centre Sheraton Montreal / Salon 3

Philosophical, Biological and Attitudinal Impacts

Chair: Randall E. Schumacker, University of North Texas

Scale-Free Genomic Measurement: Mitochondrial DNA, Nathan Markward, William P. Fisher

Designing and Validating Measures of Teacher Attitude Towards Inclusive Education (TATIE) Using an Iterative Process Model, Clarice S. Ewing, Madhabi Chatterji

Creating a Common Market for the Liberation of Literacy Capital, William P. Fisher, Jackson A. Stenner

2:15 p.m. - 3:45 p.m. Four Points by Sheraton Montreal / Mont St-Helens

An Examinee Perspective

includes:

Using Item Position and Item Difficulty to Measure Test Fatigue, Jeff Davis, Abdullah A. Ferdous

A Method for Adjusting Item P-Values for Test-Taker Motivation Using the Rasch Model, Laurie L. Davis, Michael E. Yoes

Wednesday, April 13, 2005

10:35 a.m. - 11:15 a.m. Marriott Montreal Chateau Champlain / Salle de Bal Ballroom & Foyer

Introduction to Rasch Measurement

Applying the Rasch Model: Interested Newcomers Are Invited to Talk With Authors: Trevor G. Bond, Christine M. Fox

Improving Data Collection Through Rasch Measurement: A Continuing Study of Teacher Supply and Demand, Kelly D. Bradley, Shannon O. Sampson

Functional Equivalence of English and Chinese Versions of a Cognitive Development Test for Preschoolers, Edward Wolfe, Wei He

10:35 a.m. - 12:05 p.m. Delta Centre Ville / Salon 532

Rasch Measurement: Important Aspects of the Models

Chair: Gregory E. Stone, University of Toledo

The Confusion Over Rasch and IRT or Why Don't Some Psychometricians Get Along?, Everett V. Smith, David Andrich

The Effect of Missing Data of Rating Design on Parameter Estimations Using the Many-Facet Rasch Model, Shudong Wang, Michael Young, Holly Zhang

Detecting Measurement Disturbance Effects: The Graphical Display of Item Characteristic Curves, Randall E. Schumacker, Robert E. Mount, George A. Marcoulides

Computer-Adaptive Medical Outcome Assessment: A Comparison of the Rating Scale and Successive Interval Models, Barbara G. Dodd, Karon F. Cook, Donn Godin

8:15 p.m. - 9:45 p.m. Le Centre Sheraton Montreal / Salon B

Quantitative SIGs Joint Social

Educational Statisticians, Hierarchical Linear Modeling, Survey Research in Education, Structural Equation Modeling, Advanced Studies of National Databases, Rasch Measurement

Thursday, April 14, 2005

1:00 p.m. - 5:00 p.m. Marriott Montreal Chateau Champlain / Maisonneuve, Section E & F

Professional Development Training

A Hands-on Introduction to Latent Class Models, Rasch Models and Their Extensions. Matthias von Davier.

Friday, April 15, 2005

10:35 a.m. - 12:05 p.m. Le Centre Sheraton Montreal / Salon 7

Using Rasch Measurement to Investigate Important Latent Traits

Chair: Trevor G. Bond, James Cook University

Developing an Objective Measure of Early Childhood Literacy, Stuart Luppescu, David W. Kerbow

Conditional Validation of Cognitive Structures With Rasch Measurement, Dimiter M. Dimitrov

Explanations of Translation Differences on Chinese and English Versions of a Language Test for Preschoolers, Xiaoting Huang, Edward Wolfe

Permanence of Marker Characteristics, Iasonas Lambros Lamprianou, Thekla Afantiti Lamprianou

12:25 p.m. - 1:55 p.m. Marriott Montreal Chateau Champlain / Salon 401

Engagement and Success of Freshmen

includes: A Multiple Methodological Approach to Personnel Evaluation Using Rasch Measurement Principles, IRT and Focus-Group Data, Christine M. Mills

1:15 p.m. - 1:55 p.m. Marriott Montreal Chateau Champlain / Salle de Bal Ballroom & Foyer

Rasch Analysis - Introduction to Software Packages

Rasch Measurement With *Winsteps,* Richard Smith, John M. Linacre

Metaphysics and Rasch Measurement

Campbell's Apparent Avoidance of Metaphysics

Norman Campbell (1920) remarks that:

"... one of the chief characteristics which distinguishes science from metaphysics, and the feature which makes men of science so averse from the latter, is that in science, but not in metaphysics, it is possible to obtain universal assent for conclusions, and to present results which do not lose their value because, when they are presented, they are so obvious as to be indubitable. I maintain that the results presented in this work are of that nature." (p. 10)

If "men of science find everything I have to say dull and trite and so familiar that it is not worth saying .... [p. 11:] I shall not be wholly disappointed. It will show at least that I have avoided metaphysics successfully. I am not sure that the most handsome compliment that anyone could pay my work would be to say that he knew it all before."

"But is it true that metaphysics can be avoided wholly in an attempt to probe to the foundations of science? (Now and henceforward I propose to use the word metaphysics, not as a mere term of abuse, but to denote the study which those who accept the status of metaphysician think valuable. So far as I can make out, the study consists in the investigation of reality and existence.) At some stage in our inquiry we must stop and accept judgments without argument; is it certain that these judgments will not be found to be metaphysical? Or again, are we sure that the process of reasoning by which we develop our conclusions from these fundamental judgments does not depend on the acceptance of doctrines that are distinctively metaphysical? The general opinion to-day is that science is in no way dependent on metaphysics; and the proof of independence which seems generally to be thought the most convincing is that persons holding the most diverse metaphysical views all agree in accepting the same scientific conclusion." (p. 11)

"...some men of science [e.g., E. Weichert, 1911, Phys. Zeit. 12:702] hold that science depends on the proposition, apparently metaphysical, that matter is real and exists; and many of those who are ready to assent verbally to the independence [of science from metaphysics] are apt to show great annoyance if any one dares to deny that proposition. Again many metaphysicians agree that science and metaphysics are independent only because they believe that science is not true in the same sense as is their own study; they grant its independence only at the sacrifice of its value." (p. 11)

"...we are all metaphysicians, physicists included. We are all interested in problems which the metaphysician attempts to solve. ... The world is not divided into those who do and those who do not hold metaphysical doctrines, but rather those who hold them for some reason and those who hold them for none." (p.12)

But We Are All Metaphysicians ...

Campbell interestingly takes as his criterion for successfully avoiding metaphysics as telling "men of science" only what they already know. This corresponds quite nicely with Heidegger's (1967) sense of science's mathematical metaphysics as teaching and learning through what is already known. The desire to avoid metaphysics is the same as the desire to base statements and inferences on what is known, and not on untestable or untested conjectures and speculations.

Campbell later claims to find the definition of metaphysical study obscure, saying it "consists in the investigation of reality and existence," but he is perceptive in wondering whether the process of reasoning does not depend on an at least implicit acceptance of metaphysical doctrines. He would even seem wise in recognizing that "we are all metaphysicians" whether we recognize it or not.

Problematic Positivism

As a philosopher of science, and especially as a philosopher of measurement, Campbell is unusual in accepting this point. Western philosophy, especially in matters scientific and mathematical, has a long history of positivism, which rejects metaphysics as nonsense. Positivism thus puts itself in the position of holding that it is possible to apprehend parts with no theory as to the whole to which they belong. This becomes problematic, as Campbell recognizes, as soon as any question arises as to how any instance of a species of thing is recognized. As Burtt (1954, p. 228) puts it, "even the attempt to escape metaphysics is no sooner put in the form of a proposition than it is seen to involve highly significant metaphysical propositions." And similarly, as Derrida (1978, pp. 280-1) wrote,

"There is no sense in doing without the concepts of metaphysics in order to shake metaphysics. We have no language-no syntax and no lexicon-which is foreign to this history; we can pronounce not a single destructive proposition which has not already had to slip into the form, the logic, and the implicit postulations of precisely what it seeks to contest."

Gadamer (1994, p. 187) concurs, and Burtt (1954, p. 229) accordingly asks,

"... what kind of metaphysics are you likely to cherish when you sturdily suppose yourself to be free of the abomination? Of course . . . in this case your metaphysics will be held uncritically because it is unconscious; moreover, it will be passed on to others far more readily than your other notions inasmuch as it will be propagated by insinuation rather than by direct argument."

All this notwithstanding, the philosopher Hume famously proposed that all works of metaphysics lacking quantity,

number, and matters of fact be burned (Ayer, 1952, p. 54), and his attitude is apparently shared by the vast majority of today's working scientists. Campbell, in contrast, does well to point out that questions concerning the real existence of what scientists purport to measure are metaphysical questions that often provoke great annoyance when asked of some researchers.

Intentional and Unintentional Metaphysics

And here we encounter a subtle irony. A work can be metaphysical in two ways, one intentional and the other not, as Campbell notes. Overtly metaphysical works are those that make and test explicit metaphysical postulates concerning reality and existence. Criteria for recognizing when something is real, as in able to persistently resist tests of its strength and so exist in stable states across samples, instruments, laboratories, investigators, time, space, etc., make a work metaphysical in a positive, though not positivist, sense by not attempting to avoid the inevitable.

Covertly metaphysical works are those that deny any role for metaphysics and that ignore or hide their metaphysical assumptions concerning the real existence of their objects of study. In this negative sense, a work is metaphysical to the extent that it leaves untested its assumptions concerning the nature of what is supposedly a matter of fact. The danger here is, of course, that science cannot rest content with merely confirming a researcher's prejudices and biases, but ought instead to subject these to close scrutiny and critical evaluation.

Overcoming metaphysics is then a matter of taking it up and using it (Gadamer, 1976, p. 240), since metaphysics must be presupposed even as we "get over it" (Heidegger, 1973, pp. 84-110; Gadamer, 1994, p. 164), a point missed by some commentators on the subject (Friedman, 1996). This is because, as Gadamer (1991) points out, even should we succeed in overcoming positivism's insufficient reductions and metaphysical blindness, there remains the constant danger

"of the systematic problem of philosophy itself: that the part of lived reality that can enter into the concept is always a flattened version-like every projection of a living bodily existence onto a surface. The gain in unambiguous comprehensibility and repeatable certainty is matched by a loss in stimulating multiplicity of meaning." (p. 7)

In other words, "all interpretation makes its object univocal and, by providing access to it, necessarily also obstructs access to it" (Gadamer 1991, p. 8).

Measurement and Metaphysics

In the human sciences, researchers persist in avoiding a positive use of metaphysics, and so wind up being metaphysical in the negative sense. As Michell (for instance, 2000) points out in his body of work, the most popular statistical methods in use leave untested vital hypotheses as to the quantitative structure of the variables purportedly measured by tests, surveys, and assessments. Unexamined metaphysics then remain uncritically insinuated within the measures and their rationales, with unknown potential consequences.

Rasch measurement, in contrast, especially when situated within a fully conceived and interconnected metrological system (Fisher, 2000), advances the work of overcoming metaphysics by starting from an explicit mathematical theory of what counts as real existence, applying that theory in tests of the quantitative hypothesis, and vigilantly persisting in attempting to prevent a fall into negative metaphysics (though this cannot be guaranteed) by routinely checking for adherence to the theory and by supporting decision making with replicable measures.

The Metaphysics of Meaning

Campbell says that we are all metaphysicians, but some of us take metaphysics seriously, and have good reasons for adhering to particular doctrines, while others dismiss metaphysics and wind up adhering to unarticulated doctrines for no reason. In the case of Rasch measurement, Campbell's observation is particularly apt, since the metaphysics of meaning hinge on being able to demonstrate an understanding in one's own words, or to translate representations across media (Latour, 1987). To put it more technically,

"The hallmark of a meaningless proposition is that its truth-value depends on what scale or coordinate system is employed, whereas meaningful propositions have truth-value independent of the choice of representation, within certain limits. The formal analysis of this distinction leads, in all three areas [measurement theory, geometry, and relativity], to a rather involved technical apparatus focusing upon invariance under changes of scale or changes of coordinate system" (Mundy, 1986, p. 392; also see Luce, 1978; Narens, 1981, 2002; Roberts, 1985).

By connecting numbers with invariantly additive amounts of the thing measured, Rasch models make generally accessible a heretofore only rarely attained level of meaningfulness. And meaningfulness is the most fundamental metaphysical assumption made in academic and scientific discourse. Even the most ardent deconstructionist writes, and in writing must assume that understandable meaning is communicable. This was well understood by Derrida (2003, p. 62) when he said that, in "playing with or transgressing norms," and in taking "liberties, it's always by measuring the distance from the standards I know or that I've been rigorously trained in."

Approaches to measurement that leave the quantitative hypothesis untested, in contrast, put their users in the ironic position of being more metaphysical than those who take up and use their metaphysics. In valuing quantitative methods, and in writing and publishing reports of their research, the vast majority of investigators in the human sciences are transgressing their own metaphysical doctrines concerning the transparency and generalizability of their results. But in opposition to Derrida's deconstructionist strategy, they do so in a way

that blindly does not measure their distance from the measure-value standards they take for granted in any grocery store purchasing decision.

A Metaphysical Faith

Rasch measurement practitioners are sometimes criticized for being zealous adherents to what is said to be but one metaphysical faith among many equal faiths. But it seems to me that it is much more a matter of deciding where one's faith is best invested, in empirically tested and theoretically informed, generalizable, invariant meaningful mathematical structures, or in blindly assumed, atheoretical, and ungeneralizable scale-dependent scores? To paraphrase Derrida (1989, p. 218), as soon as you give up metaphysics, or the word metaphysics, and you think you have overcome it, defeated it, what happens is not something new or beyond metaphysics. Instead, what happens is that some old hidden metaphysics persists under the cover of another name, such as this new method or that collection of test items, or what have you, and the same old metaphysics goes on dominating the research in an implicit or dogmatic way. When you want to make this implicit metaphysics as clear as possible by establishing the invariance of truth-values over representations, you have to accept being a metaphysician and go on philosophizing.

Successful Scientific Metaphysics

In a field like ours, where the objects of investigation are still so early in their conceptual, gestational, and maturational processes that successful birthings are far from certainties, we must be as vigilant as possible in providing the nurturing environment needed to bring them to life. My bets are on the methods that take up and use the metaphysical assumptions that have proven foundational to the history of science. The ancient Greek metaphysics of learning through what is already known (sign and symbol systems), and what is learned in this way (writable meaning), comprise "the fundamental presupposition of all 'academic' work" and "of the knowledge of things" (Heidegger, 1967, pp. 75, 76). To be an academic is to accept in practice, if not in theory, "that we today, after two thousand years, are still not through with this academic work and never will be so long as we take ourselves seriously" (Heidegger, 1967, p. 76). The integration of theory and practice would seem to require further attention to our metaphysics, and efforts aimed toward achieving Campbell's goals of telling ourselves only what we already know, of not being annoyed by questions as to the reality of our objects of study, and of being prepared with good answers to those questions.

William P. Fisher, Jr.

Ayer, A. J. (1952). Language, Truth and Logic (2d). New York, New York: Dover.

Burtt, E. A. (1954/1932). The metaphysical foundations of modern physical science (Rev. ed.) [First edition published in 1924]. Garden City, New York: Doubleday Anchor.

Campbell, N. R. (1920). Physics, the elements. Cambridge, England: Cambridge University Press.

Derrida, J. (1978). Structure, sign and play in the discourse of the human sciences. In Writing and difference (pp. 278-93). Chicago, Illinois: University of Chicago Press.

Derrida, J. (1989). On colleges and philosophy: An interview conducted by Geoffrey Bennington. In L. Appignanesi (Ed.), Postmodernism: ICA documents (pp. 209-28). London, England: Free Association Books.

Derrida, J. (2003). Interview on writing. In G. A. Olson & L. Worsham (Eds.), Critical intellectuals on writing (pp. 61-9). Albany, New York: State University of New York Press.

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

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

Friedman, M. (1996). Overcoming metaphysics: Carnap and Heidegger. In R. N. Giere & A. W. Richardson (Eds.), Minnesota Studies in the Philosophy of Science. Vol. XVI: Origins of logical empiricism (pp. 45-79). Minneapolis, Minnesota: University of Minnesota Press.

Gadamer, H.-G. (1976). Philosophical hermeneutics (D. E. Linge, Trans.). Berkeley: University of California Press.

Gadamer, H.-G. (1991). Plato's dialectical ethics: Phenomenological interpretations relating to the Philebus (R. M. Wallace, Trans.). New Haven, Connecticut: Yale University Press.

Gadamer, H.-G. (1994). Heidegger's ways (D. J. Schmidt, Ed.) (J. W. Stanley, Trans.). SUNY Series in Contemporary Continental Philosophy. Albany, New York: SUNY Press.

Heidegger, M. (1967). What is a thing? (W. B. Barton, Jr. & V. Deutsch, Trans.). South Bend, Indiana: Regnery/Gateway.

Heidegger, M. (1973). The end of philosophy. (J. Stambaugh, Trans.). New York: Harper & Row.

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

Luce, R. D. (1978). Dimensionally invariant numerical laws correspond to meaningful qualitative relations. Philosophy of Science, 45, 1-16.

Michell, J. (2000, October). Normal science, pathological science and psychometrics. Theory & Psychology, 10(5), 639-667.

Mundy, B. (1986). On the general theory of meaningful representation. Synthese, 67, 391-437.

Narens, L. (1981). A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. Theory and Decision: An International Journal for Philosophy and Methodology of the Social Sciences, 13, 1-70.

Narens, L. (2002). Theories of meaningfulness (S. W. Link & J. T. Townsend, Eds.). Scientific Psychology Series. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Roberts, F. S. (1985). Applications of the theory of meaningfulness to psychology. Journal of Mathematical Psychology, 29, 311-32.

"Probable impossibilities are to be preferred to improbable possibilities"

Aristotle, Poetics 24, 1460a.

Journal of Applied Measurement
Volume 6, Number 1. Spring 2005

Rasch Analysis of Inattentive, Hyperactive, and Impulse Behavior in Young Children and the Link with Academic Achievement. *Christine Merrell and Peter Tymms, 1-18*

Measuring Statistical Literacy. Rosemary Callingham and Jane Watson, 19-47

Expected Linking Error Resulting from Item Parameter Drift among Common Items on Rasch Calibrated Tests. *G. Edward Miller, Paul Randall Gesn, and Ourania Rotou, 48-56*

Measuring College Sailing Teams Ability: An Application of the Many-Facet Rasch Model to Ordinal Data. *William Steve Lang and Judy R. Wilkerson, 57-70*

On the Lack of Comonotonicity between Likert Scores and Rasch-Based Measures. *Lucio Bertoli-Barsotti, 71-79*

An Analysis of Dimensionality using Factor Analysis (True-Score Theory) and Rasch Measurement: What is the Difference? Which Method is Better? *Russell F. Waugh and Elaine Chapman, 80-99*

Does Data Rounding-Off Influence Reproducibility Index Estimates? *Bruno Giraudeau, Philippe Ravaud, and Jean-Yves Mary, 100-108*

Understanding Rasch Measurement: Computer Adaptive Testing. *Richard C. Gershon, 109-127*

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

What Good are Statistics that Don't Generalize?

This question is the title of a paper by David Williamson Shaffer and Ronald C. Serlin in *Educational Researcher,* 2004, 33:9, 14-23. Their answer is: "Use ISSA!"

"ISSA [Intra-sample Statistical Analysis], then, extends the concept of verbal analysis to encompass qualitative data of any kind, including observations as well as interviews, task analyses as well as think-aloud protocols, video and field notes as well as audio transcripts. More importantly, ISSA provides a theoretical justification for the use of statistical analyses to support qualitative inference – and thus an occasion to reexamine the assumptions of quantitative and qualitative research traditions themselves." (p. 16).

ISSA parallels Rasch in its aim of progressing from observed manifest qualities, to abstract generalized quantities, to inferred latent qualities. Indeed, perhaps Rasch is an instance of ISSA.

But, as William Fisher remarks:
"Though I applaud its overall purpose and tone, it is too bad the authors, reviewers, and editors responsible for the Shaffer & Serlin (2004) article in the December *Educational Researcher* did not connect the qualitative data pattern shown in their Figure 1 with measurement theory. I find myself in a permanently perplexed state as a result of this near-constant hammering of the general unawareness that scaling and calibration inherently involve qualitative pattern-finding and pattern-imposing. Statisticians in general and the ones in this article seem unable to see past their descriptive orientation."

Student	Problem	Related problem to personal experience	Answered correctly	Related problem to personal experience	Answered correctly
A	1	✓	✓	2	2
	2	✓	✓		
	3	X	X		
	4	X	X		
B	1	X	X	2	2
	2	X	X		
	3	✓	✓		
	4	✓	✓		
C	1	X	X	3	2
	2	✓	✓		
	3	✓	✓		
	4	✓	X		
D	1	✓	✓	3	1
	2	✓	X		
	3	✓	X		
	4	X	X		

Here is Shaffer and Serlin's *Figure 1:* "A hypothetical subset of data on answers to post test questions shows how patterns observed can be obscured in an analysis of aggregated data" (p. 19). This data would immediately support two standard Rasch analyses: (a) measures for persons and items when "related to personal experience", and (b) when "not related". Cross-plots of the person measures would reveal **who** is impacted, and cross-plots of the item difficulties would reveal **what** is impacted. But Shaffer and Serlin do not attempt Rasch analysis of their data, instead their focus is on weakening, rather than strengthening, the axiomatic basis of the numbers they produce.

Item Discrimination, Test Optimization and W. E. Deming

Take a look at the most discriminating item on your Test. This item operationalizes your *best effort* at separating your high performers from your low performers. From the perspective of this item, the distance on the latent variable between the high and low performers is greater than it is for any other item in your Test. ***Wonderful!?***

Statistician Wayne Edwards Deming observed this type of optimization in many industrial processes. Here is a composite of some of his examples: Several machine tools were manufacturing the same component. Different operators employed different tactics for maximizing usable output. Consequently, some of those machine tools were set to tighter tolerances than specified, in order to minimize out-of-"official"-tolerance components. Some operators set their machines within "official" tolerance limits, but deliberately made components toward the larger end of the tolerance interval, so that components could be easily remachined smaller if discovered to be out of tolerance. In fact, every manufactured component represented a personal "best effort" by a machine-tool operator. But W. E. Deming perceived that **"We are being ruined by our best efforts"** (Neave, 1992).

Optimizing each part does not necessarily optimize the whole. For instance, separately optimizing the performance of each member of a basketball team may not optimize team performance. Separately optimizing each component of an audio amplifier may not optimize audio output quality, indeed with some designs may worsen it. Separately optimizing the skills of each musician may not make the orchestra perform better.

The problem with "best efforts" is that they tend to focus on the immediate situation, ignoring the larger context. Those machine operators were given the specifications for the part they were producing, but had no idea how each of their individual approaches impacted the overall quality of the final product. In fact, the highest quality final product was produced by using the widest allowable tolerance range (which was wider than the overly-cautious design engineers originally specified), and setting the machine tool to work in the center of it. This also reduced rejection rates, remachining and improved component interchangeability.

The same is true of Test items. Allowing excessive variation in item discrimination may optimize individual items, but that variation degrades the meaning and utility of the Test as a whole. So how do we know when an optimizing tactic will work? **"Only theory can help us figure out what's right and what's wrong"** - Deming again. The Rasch model tells us to aim at the center of the discrimination range, and permits us some, but not too much, variation (RMT 14:3, 743).

Neave H.R. (1992) The Deming Dimension. Knoxville, TN: SPC Press.

Erling B. Andersen

Erling Andersen, Professor in the Economics Institute, University of Copenhagen, and student of Georg Rasch, died on 18 September 2004, at 64 years old. Here is part of his personal recollection of Henri Caussinus:

"When writing my book on categorical data, published in 1980, I included a chapter on two-way contingency tables with dependencies. From my teacher and predecessor Georg Rasch I have learned about the model, later to be called RC-association models by Leo Goodman, but I still looked around to see if somebody had not addressed this problem. In my search I became aware of the paper by Henri Caussinus in the *Annales de la Faculté des Sciences de l'Université de Toulouse* [1]. I shall not claim that I understood all 111 pages in French. But I immediately saw the importance of Henri's work, and (as anybody can check,) he is duly quoted in my 1980 book.

"It was some years later that I actually met Henri. Actually at a COMPSTAT meeting in Copenhagen. I did not know what to expect. But it so happened, that I had just started playing the French Horn, and persuaded my teacher, solo player in the Royal Danish Orchestra, to give a concert in one of the churches in Copenhagen, together with several of his colleagues in the Royal Orchestra. I remember that my wife Ellen and Henri and me sat in one of the front boxes of the church. Henri really enjoyed the music, and he was much pleased, when we after the concert got a little chat with the royal musicians.

"Some years later I discovered - maybe a bit later than several of my Dutch friends and colleagues - the importance of correspondence analysis and its relationship to the RC-association model as formulated by Georg Rasch and Leo Goodman. I decided, therefore, to spend a couple of months with professor Escoufier in Montpellier, especially to unravel the mystery of MCA = Multiple Correspondence Analysis. During my stay in Montpellier, I was invited to come to Toulouse for two days. It became two of those days that will always stay in my mind as a combination of many long talks of what had happened to the theory of contingency tables for the last 20 or so years, and of other long talks of what has happened to our children and what we enjoyed and expected of our grandchildren."

[1] Erling B. Andersen. The Statistical Analysis of Categorical Data. Springer-Verlag, Berlin Heidelberg New York, third edition, 1994.

[2] Henri Caussinus. Contribution à l'analyse statistique des tableaux de corrélation. Annales de la Faculté des Sciences de l'Université de Toulouse, 29 (année 1965):77–183, 1966.

Excerpted from www.lsp.ups-tlse.fr
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
Toulouse, France.