Research Problems– Rasch Solutions

Donna Surges Tatum, Ph.D.

DECISION MAKING

We conduct research because we have questions about how to react to a given situation. The time, energy and money invested in the research and the effects of decisions require confidence in the research process. Unfortunately the complete information contained in the data does not always see the light of day. This is because traditional data analysis techniques do not access the subtleties and complexities inherent in most research situations.

We know that there are problems we should deal with when analyzing data. But because we do not know how to do so, we do the best we can with what we are used to. Today techniques enable us to address these problems directly and efficiently, instead of having nightmares about them.

RATING SCALES

RAW RATING SCALES DO NOT HAVE A UNIFORM, LINEAR STRUCTURE

Rating scales are one of the most commonly used research tools. Surveys, evaluation instruments, and psychological tests depend on ratings. Standard analyses treat these ratings as if the choices were evenly spaced steps equally separated. This is not the case.

Research shows that the spacing around rating choices are not equal. Many raters have a tendency to group their choices around the middle of the scale values. The end categories are further from the points next to them than the other categories are from each other, because some raters do not like to make extreme judgments.

Instead of the intention that each category on the scale be evenly spaced:

1 2 3 4 5 6

Reality is messier:

1		2	3	4	5	6
terrible	poor		fair	good	very good	excellent

RAW SCORES ARE NOT SUITABLE FOR ADDING AND AVERAGING

ITEMS

ALL ITEMS ARE NOT EQUAL

When surveying for such things as attitudes, speech confidence, or speaking ability, the items used are not all at the same point on the scale. Some items demand a more intense attitude than others, or a greater level of ability.

It is easier for students to agree that they are more comfortable preparing a speech than that they enjoy giving speeches. It is easier for them to demonstrate knowledge of their topic than to have good gestures.

Indeed, it would not be useful if all items did measure at the same point on the scale. That would not allow us to discover the structure of the variable. Important information is contained in the differences between elements, the difference between hard and easy items. Understanding the hierarchal structure of the items improves information for decision-making.

ITEMS MUST BE PROVEN VALID AND RELIABLE

Items must also be examined to determine whether they all relate to the same variable, or whether there are different subscales. The items must behave in a predictable manner. When some items are misunderstood by those that use the rating form, we must discover this. We must find out whether our items fit the theoretical construct we intend — the idea which motivates our research.

EXAMINE ITEMS FOR ORDER OF DIFFICULTY AS WELL AS VALIDITY

Donna Surges Tatum earned her B.A. and M.A. from Purdue University in Communication with an emphasis on Persuasion and Organizational Communication. She moved to Chicago upon graduation to join the "real" world. For seven years she worked in advertising and marketing until she realized that she was only in an alternate reality. She became a consultant and returned to academia, teaching at Roosevelt University. She was Director of the Communication Studies Department from 1986 to 1989.

Donna received her Ph.D. in 1991 from MESA at the University of Chicago. She has been teaching since 1990 in the Graham School of General Studies at the University of Chicago. It must mean something (she's not sure what) that she teaches the two courses most hated by most people: Public Speaking and Statistics.

In 1991 Donna started Meaningful Measurement, a consulting consortium for communication training, organizational development, market research, and educational assessment. Her leisure activities are swimming, yoga, and reading mystery novels.

e-mail: surgstatum@aol.com

RATERS

ALL RATERS ARE NOT EQUAL — THEY ARE INDIVIDUAL IN THE WAY THEY JUDGE A SITUATION

Raters are a crucial element in many research projects. We know from Communication and Psychology theory that we each live in our own perceptual world, and attend to our own things. One person will react more to how a speech is organized than how it is delivered. Another may be the opposite.

No matter how hard we try to train raters, we will never achieve the ideal in which all raters are the same. Instead of a false assumption of sameness, we must address the issue of differences. In fact, the real differences between raters is important additional information.

But different raters have different levels of severity when judging an event, thus we cannot take their raw scores and add them to come up with an objective measure. One rater's "3" may be worth more than another rater's "4" because that first rater is consistently more critical in her judgments. Once again we see that we cannot use the raw scores for mathematical functions.

RATERS MUST BE CONSISTENT IN THEIR JUDGMENTS

We hope that our raters are well-trained and well-behaved. But if a rater is inconsistent in judgment, then we must be able to detect who is or is not providing consistent evaluations. Otherwise we will have no basis upon which to make comparisons.

RESULTS

AN AVERAGE OR PERCENTAGE IS NOT A MEASURE

When results are given in terms of raw scores with averages or percentages, they are descriptive of one-time events. The results are not true measures because they can not be used to perform arithmetic functions such as addition, subtraction, and multiplication.

One of the fundamental errors made in research is to use scores to perform a function for which they are not equipped — to measure instead of describe. This is like using a "rubber ruler;" there is no consistency or comparability between persons, items, or groups. Scores describe a one-time event, after which the rubber ruler has to be thrown away because it is of no further use. It is not a calibrated ruler of units with fixed intervals. There is no common frame of reference with standardized measures. Subsequent research will be "measured" with another rubber ruler that is not really the same thing, even though the appearance is the same. This leads to fuzzy descriptions instead of facts of measurement.

DIRECT COMPARISONS REQUIRE A STRAIGHT LINE

Without a straight line marked in equal intervals, direct comparisons lack precision and accuracy. Tracking products over time, from group to group, or in field tests can be tedious, difficult, and imprecise. If a calibrated ruler is used to measure instead of a rubber ruler, then pictures and maps can be drawn to show the results. A well-drawn picture is worth a thousand numbers. It creates perspective.

A STABLE FRAME OF REFERENCE MUST BE CREATED AND MAINTAINED TO MAKE MEANING OUT OF DATA

SOLUTION

Many years of careful research produced a scientific method based on the Rasch Model. This system for research and data analysis is Objective Measurement. In 1953 Georg Rasch, a Danish mathematician, was hired by the Danish government to develop achievement tests to place army recruits. He discovered a mathematical model that was completely different from any used previously for this type of data analysis. In 1960 Rasch came to the University of Chicago for a year where he met Benjamin D. Wright. Professor Wright, a psychologist who originally trained as a physicist, saw the implications of this method. In 1963 he founded the MESA Psychometric Laboratory at the University of Chicago where he and his colleagues refined and extended the Rasch model. In the process they revolutionized social science research.

METHOD IN BRIEF

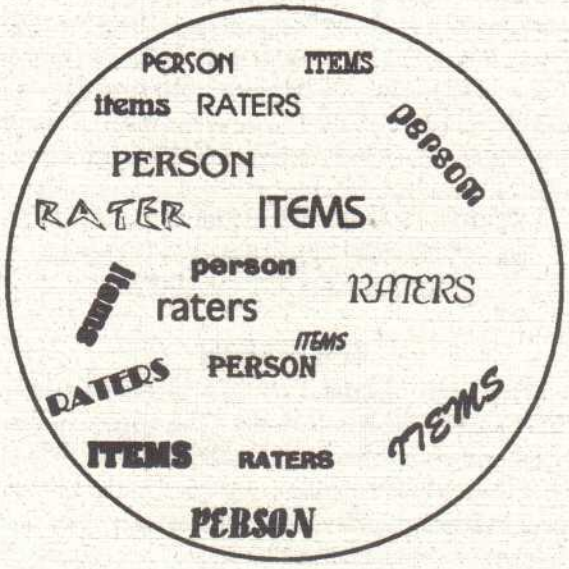
This is a brief explanation of the concepts inherent to understanding Objective Measurement. This unique approach to rater-mediated evaluations provides the most objective means for assessment yet discovered.

The Research Situation:

A traditional analysis of raw scores is primarily descriptive. It gives us a simple snapshot of the research situation. It portrays a specific group of people using a particular set of test items at a given time. All the elements are inextricably bound together. Raw scores are not linear, and do not have the mathematical properties of true measurement.

Social scientists take a snapshot of the research situation as represented by the circle below. They or others replicate the snapshot and then compare snapshots. However, these circles are not directly comparable. Each one is unique unto itself. Each circle reflects a particular, discrete situation. Averages, percentages, or percentiles based on raw scores are sample dependent, and can only represent what is happening

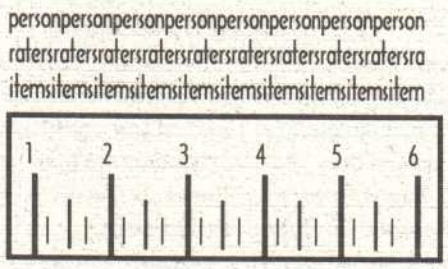
in **that** circle with **those** elements at **that** time. The results are not a measure that transcends from the particular to the general.



Measured Elements

When raw scores are conditioned using Objective Measurement techniques, something wondrously useful occurs. The strands in the analysis are disentangled from each other, and smoothed out into straight lines. They are calibrated into common units, providing context-free rulers that are able to measure at any time and any place. These results are precise reproducible measurement instead of fuzzy idiosyncratic descriptions of statistics.

Investigation is now possible in a manner that conforms to scientific principles. Instruments are constructed and calibrated to produce generalizable results. Each element can be examined separately, allowing us to delve into the data in a far deeper way than has been possible with traditional methods. We discover information heretofore unavailable.



This is it in a nutshell:

Observational statistics like raw scores and ratings describe a one-time event with all elements interwoven. Objective Measurement gives us straight lines, precise measures, and separated elements that remain stable across time and sample.