

Some Insights into Objective Measurement

David Andrich, Ph.D.

There are many aspects of my studies in quantitative methods with Georg Rasch that have been important in my work, but I will focus here on two related insights that were the most telling in affecting my outlook on measurement.

I was a student at the University of Chicago in 1971–1973, at a time when the Department was extremely exciting. Among his many favors, Ben Wright did his biggest one by introducing me to Georg Rasch. On completing my Ph.D., I visited Rasch in Denmark and arranged for him to be a visiting professor in the Departments of Education and Mathematics at the University of Western Australia in 1974. I spent many hours during the day with Georg, and my wife Joan and I enjoyed the company of Georg and his wife Nille during the evenings and weekends for seven months. We then repeated the pleasure in Denmark in 1977 for another five months.

In studying general quantitative methods in the social sciences, I had learned a whole range of techniques and skills for using models and analyzing data. In addition, however, I learned the implied general philosophical position behind these studies, namely, that the task is to find a model that accounts for the data. One could debate this position in general, but in the case of measurement, I realized through the work with Rasch that the case for his class of models does not depend on modeling any particular data. This was a very important shift in perspective for me, and I believe that where there is controversy in the use of Rasch models, it is where people consider that the choice of one model or another rests essentially on how the models account for data. The Rasch class of models are justified as expressions of the requirements of measurements; they are not justified as descriptions of data. Although it now seems obvious, at the time it seemed a very important insight to me.

The case for the model rests on the requirements of measurement, and if data are to be transformed to measurements, then they must be valid expressions of the construct in all the traditional senses, and in addition, need to meet the requirements of the Rasch class of models. In the special case of dichotomous responses, the discrimination at the difficulties (item thresholds) has to equal. To estimate the discriminations destroys the requirement of invariance of person ability estimates in the model, and if items have different discriminations, then item difficulties can take different orders depending on the distribution of the persons. Of course, it is an empirical question as to the degree that real data show equal discriminations, and many data sets will not immedi-

ately have equivalent discriminations. That is why it is tempting to try to estimate the discriminations. It is not surprising that data sets which have been collected without an eye to these requirements do not follow a model with the equal discriminations criterion — indeed it is surprising how many data sets do follow the models sufficiently well to be informative.

My second insight came with the resolution of the coefficients in Rasch's form of the multcategory model. Rasch generalized his dichotomous model to one for many categories as a multidimensional model and then specialized it to the case of a single location parameter for the persons and items. In the process he had a coefficient and a scoring function for each category.

These coefficients and scoring functions were difficult to make sense of in any concrete way. I constructed the model for the response of a single person to a single item beginning with the simple model for dichotomous responses at each threshold of a multcategory item, and that gave me the integer scoring function and the resolution of the category coefficients into the sums of successive thresholds. I did this while Rasch was in Perth in 1974. However, more data sets than not showed reversed thresholds in their estimates, which was inconsistent with the construction of the model. While in Copenhagen, in 1977, Erling Andersen showed me a prepublication copy of a paper to appear in *Psychometrika*, in which he showed that the scoring functions had to have a constraint. My integer scoring functions had such a constraint, which confirmed to me that I was on the right track, but as I indicated, more data sets than not showed a problem with the estimates. My insight came in realizing that when the threshold estimates were reversed, this was not a problem with the model, but with the data. In particular, if discriminations at the thresholds were not equal, then it was possible, as in the dichotomous case, to get any ordering of the thresholds, depending on the distribution of the persons. Again, after the formalization, the result seemed obvious and a simple generalization of the dichotomous case. However, at the time, it revealed the level of resistance in my mind in taking seriously that the case for the model rests on criteria independent of the data, and not in modeling data. Because of our traditional studies in quantitative methods, it is much easier to think that the model should describe whatever data are at hand, and it is difficult to maintain in our thinking that the case for the Rasch models become independent of data. It is also difficult to resist the temptation to use other models to model the data, rather than to examine the data to see how and why they violate the requirements of measurement.