A TAXONOMY OF ITEM RESPONSE MODELS

DAVID THISSEN
UNIVERSITY OF KANSAS

LYNNE STEINBERG
EDUCATIONAL TESTING SERVICE

A number of models for categorical item response data have been proposed in recent years. The models appear to be quite different. However, they may usefully be organized as members of only three distinct classes, within which the models are distinguished only by assumptions and constraints on their parameters. "Difference models" are appropriate for ordered responses, "divide-by-total" models may be used for either ordered or nominal responses, and "left-side added" models are used for multiple-choice responses with guessing. The details of the taxonomy and the models are described in this paper.

Key words: item response theory, test theory, logistic models.

Over the past few years, a growing (and bewildering) array of models has been described and proposed for use in item response theory. In the beginning, there were only a few models, named either after their originator, as with the Rasch (1960) model, or after the mathematical function itself, as in the "normal ogive" model. But naming models after persons produces difficulties: it is sometimes a problem determining who the right person is, and some persons (e.g., Samejima) have invented so many models that unique nomenclature becomes problematic. Another difficulty arises when considering how much a model may be modified before a change of name is warranted. The "Rasch model" is a case in point: Rasch (1960) did not explicitly consider the population distribution, so now there is some disagreement about whether "equal-slope logistics" with some consideration of the population distribution is the "Rasch" model or not.

Naming the models after their constituent curves is not much more convenient, because imposing constraints on the parameters of a model typically makes a new model, although it still uses the same mathematical function. The most recent wave of nomenclature employs the proposed function of the model; such models as the "rating scale model" (Andrich, 1978), the "partial credit model" (Masters, 1982), and the "multiple choice model" (Thissen & Steinberg, 1984) are examples. This is satisfactory until there get to be two "rating scale models," for instance; then modification is required. In addition, naming models in terms of their proposed function may have the effect of imposing limits regarding the appropriate use of a model by implying that they are useful only in their named context.

The purpose of this paper is to introduce a taxonomy of some extant item response models, in the hope of clarifying some of the very close interrelationships among them, and suggesting new models to fill in obvious gaps in the current set. The taxonomy is presented in Figure 1; the reasoning behind it is the subject of the remainder of the paper.

The present study was supported in part by two postdoctoral fellowships awarded to Lynne Steinberg: an Educational Testing Service Postdoctoral Fellowship at ETS, Princeton, NJ and an NIMH Individual National Research Service Award at Stanford University, Stanford, CA. Helpful comments by the editor and three anonymous reviewers are gratefully acknowledged.

Requests for reprints should be sent to David Thissen, Department of Psychology, University of Kansas, Lawrence, KS 66045.
TAXONOMY

BINARY MODELS
- Normal Ogive
  (Lawley, Tucker, Lord)
- Logistic: 1PL, 2PL
  (Rasch, Birnbaum)
- Splines
  (Winsberg)
- Others?

DIFFERENCE MODELS
- Graded
  (Samejima)

DIVIDE-BY-TOTAL MODELS
- Partial Credit
  (Masters)
- Rating Scale
  (Andrich)
- Others
  (Masters & Wright)
- Nominal
  (Bock)

LEFT-SIDE ADDED MODELS
- 3PL
  (Birnbaum)
- "2 parameter"
  (Choppin)

LEFT-SIDE ADDED DIVIDE-BY-TOTAL MODELS
- Multiple Choice
  (Samejima)
- Model 6
  (Sympson)
- Multiple Choice
  (Thissen & Steinberg)

FIGURE 1.
The Taxonomy.