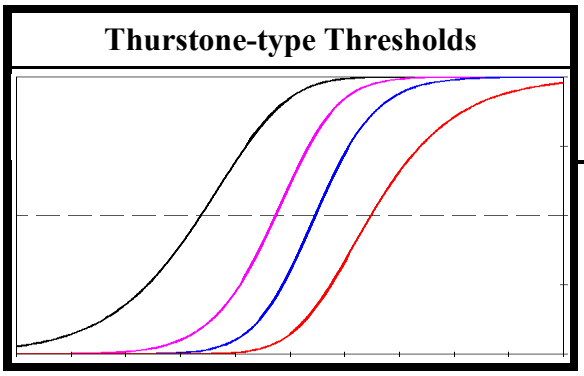


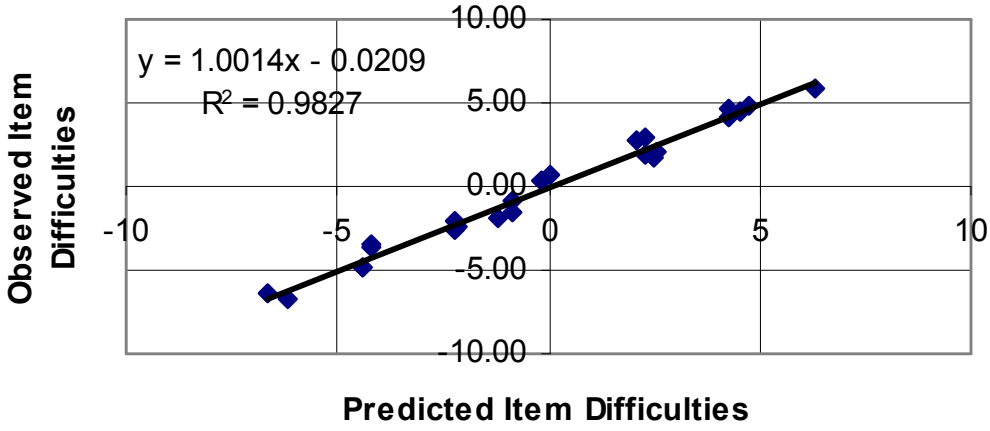
RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 17 No. 3 Winter 2003 ISSN 1051-0796

Item Specification vs. Item Banking

Our thesis is simple and straightforward. It is not necessary to have a bank of items for measuring a construct when we possess an algorithm for writing an item at any desired level of difficulty. The algorithm is the *key to the bank*, so to speak. If one has the key, the bank is open.



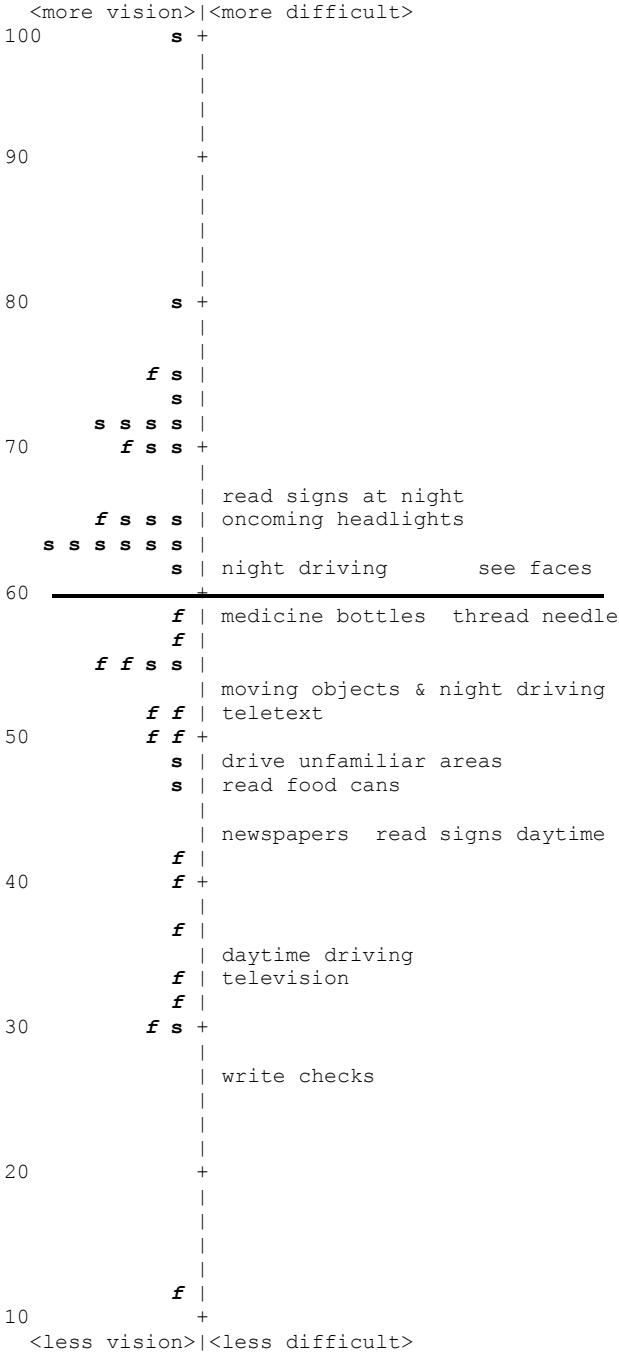
Bruce Choppin (1968) was an early Rasch pioneer who promoted item bank development. Items representative of the variable of interest are banked and selected for use as required. Leveled paper-pencil tests can be quickly assembled from the bank of items based on their associated item calibrations and item use histories. Also, computer based adaptive tests can be assembled electronically and targeted to each examinee. As useful as item banking has proven to be it is possible to move beyond the banking of individual items and their associated item statistics.

When enough is known about what causes item difficulty a specification equation can be written that yields a theory based item calibration for any item the computer software designs. An item's calibration is seen to be the consequence of decisions the computer software makes in constructing the item. This process mimics the steps a human item writer takes in constructing an item, albeit, with more control over the causal recipe for item difficulty. A thesis of this paper is that when asserting that a measure possesses construct validity there is no better evidence than demonstrated experimental control over the causes of item difficulty.

A measurement instrument embodies a construct theory; a story about what it means to move up and down a scale (Stenner, Smith & Burdick, 1983). Such a theory should be vigorously tested. In a demonstration of these methods Stone (2002) theorized that the difficulty of short term memory and attention items (Knox Cube Test) was caused by (1) number of taps, (2) number of reverses in the direction of the tapping pattern and (3) total distance in taps for the pattern. This theory was tested by regressing the observed item difficulties on the above mentioned three variables. The Figure plots the correspondence between predicted (theoretical) item difficulties and observed item difficulties. Ninety-eight percent (98%) of the variation in observed item difficulties was explained by number of

Table of Contents

Maps for Diagnosis and Prediction



The "bilateral cataract" patients are waiting for first (**f**) or second (**s**) eye cataract surgery. The line at 60 units roughly divides "**f"** from "**s"** patients. It indicates an expected vision level after first surgery. Vision may be good enough for day-time driving, but not for night-time.

Pesudovs K., Garamendi E., Keeves J.P., & Elliott D.B. (2003) The Activities of Daily Vision Scale for cataract surgery outcomes: Re-evaluating validity with Rasch analysis. *Investigative Ophthalmology & Visual Science,* 44, 7, 2892-2899.