Psychology Science Quarterly, Volume 51, 2009 (2), pp. 181 - 194

Rasch and pseudo-Rasch models: suitableness for practical test applications[1]

HARTMANN H. SCHEIBLECHNER[2]

Abstract

The Rasch model has been suggested for psychological test data (subjects × items) for various scales of measurement. It is defined to be specifically objective. If the data are dichotomous, the use of the dichotomous model of Rasch for psychological test construction is almost inevitable. The two- and three-parameter logistic models of Birnbaum and further models with additional parameters are not always identifiable. The linear logistic model is useful for the construction of item pools. For polytomous graded response data, there are useful models (Samejima, 1969; Tutz, 1990; and again by Rasch, cf. Fischer, 1974, or Kubinger, 1989) which, however, are not specifically objective. The partial credit model (Masters, 1982) is not meaningful in a measurement theory sense. For polytomous nominal data, the multicategorical Rasch model is much too rarely applied. There are limited possibilities for locally dependent data. The mixed Rasch model is not a true Rasch model, but useful for model controls and heuristic purposes. The models for frequency data and continuous data are not discussed here. The nonparametric ISOP-models are "sample independent" (ordinally specifically objective) models for (up to 3 dependent) graded responses providing ordinal scales or interval scales for subject-, item- and response-scale-parameters. The true achievement of sample-independent Rasch models is an extraordinary generalizeability of psychological assessment procedures.

Key words: specific objectivity; measurement structures; graded responses; local dependence; generalized assessment procedures

---

[1] Invited speech at the conference of „Differential Psychology, Personality Psychology, and Psychodiagnostics" of the German Society of Psychology, Vienna 2007

[2] Correspondence should be adressed to Hartmann Scheiblechner, PhD, Philipps-Universität Marburg, Biegenstraße 10/12, 35037 Marburg, Germany; email: scheible@mailer.uni-marburg.de

1. Basic concepts

Any kind of measurement requires comparison. There are no directly available observable outcomes of the contact of the 'intelligences' of two subjects (or of the 'difficulties' of two items). We can only observe the reactions of subjects to items which do depend on both simultaneously. We need instruments to compare latent dimensions. The items are the instruments used for the comparison of subjects and vice versa. The difficulty in (psychological) measurement is to achieve comparisons on a set of subjects (or a set of items) in regard to a specific latent dimension which do not depend on the choice of instruments used for the comparison.

Definition: Let <A×Q, P> be a '*frame of reference*' (Rasch, 1961) or a '*(probabilistic) instrumental paired comparison system*' (Irtel, 1987) where $A=\{a,b,c,\ldots\}$ is a set of subjects and $Q=\{x,y,z,\ldots\}$ is a set of items and $P(t;a,x)=P\left[T\leq t|a,x\right]$ is a family of cumulative distribution functions (c.d.f.s) defined on the set $T=\{t,t',t'',\ldots\}$ indexed by $A\times Q$.

Definition: An instrumental paired comparison system is a *Rasch model* if it is *specifically objective*, i.e. if there exist comparison functions on the sets $A\times A$ and $Q\times Q$ which do not depend on the parameters of the respective other set, i.e.

$v(a,b)$ independent of $\{x,y,z,\ldots\}$ and
$w(x,y)$ *independent of* $\{a,b,c,\ldots\}$

Pseudo-Rasch models are models which call themselves Rasch models without being specifically objective.

The present section defines the basic problem. The subsections of Section 2 are organized along the lines of Fischer's (1974) proof of the uniqueness of the specific objectivity of the dichotomous Rasch model among dichotomous models. Each assumption will be examined with respect to usefulness and necessity. Fischer's list of necessary and sufficient conditions ensures that no important aspect is overlooked. Section 3 illustrates the true advantages of Rasch models for theory and practice. Section 4 presents nonparametric Rasch models together with an empirical application. The discussion in Section 5 summarizes the results.

2. Dichotomous indicators

Qualitative dichotomous observations are the most elementary outcomes of instrumental comparisons. Is Birnbaum's two-parameter logistic model (2PL) an improvement over the dichotomous Rasch model? The response function of Birnbaum's model is given by

$$p(x_{vi} \mid v,i) = \frac{\exp\left[x_{vi}\alpha_i\left(\xi_v - \sigma_i\right)\right]}{1+\exp\left[\alpha_i\left(\xi_v - \sigma_i\right)\right]} \qquad x_{vi} = \begin{cases} 0 & wrong \\ 1 & right \end{cases} \qquad \alpha_i = 1: \quad Rasch\_model$$

with the sufficient statistic

$x_{v.} = \sum_i x_{vi}\alpha_i$ ($v = 1,2,\ldots,n$, the marginal sum) for the subject parameter (and

$x_{.i} = \sum_v x_{vi}\alpha_i$ $\;i = 1,2,\ldots,k$, for the item difficulty parameter).

First, the sufficient statistic is not really a statistic because it depends on the unknown discrimination parameters α_i . Second, if the item discrimination parameters α_i are continuous real numbers, then the probability that two discriminations are equal is zero:

$$p\left(\alpha_i = \alpha_j, i \neq j\right) = 0 \qquad a < \alpha_i, \alpha_j < b\,.$$

If k is a third item $i \neq j \neq k$ then the probability that two of the discriminations are equal or that a parameter is equal to the sum of the two others is zero; and so on for finite numbers k of items. Therefore, for 2^k different response vectors there are 2^k different sums and different values of the subject parameter. A unique sufficient statistic corresponds to each response vector and vice versa; the statistic is *trivially sufficient*. The probability of the only possible response vector given a marginal sum is 1. There are $2k$-2 item parameters and 2^k subject parameters in the model and 2^k-1 degrees of freedom in the data. The parameters are not identified. To make them identifiable, additional assumptions must be added to the model, e.g. about the distribution of subject parameters in the marginal maximum likelihood method – which, however, is diametrically opposed to specific objectivity. The problem persists even if the discrimination parameters are assumed to be known.

The three-parameter logistic model (3PL), with an additional guessing parameter, and further models (Fischer & Molenaar, 1995; van der Linden & Hambleton, 1997) with additional parameters – intended as improvements of the dichotomous Rasch model – make the models more flexible and may increase the *descriptive* power of the models but are not desirable as measurement models. Psychometrics – the development of new measurement procedures for latent dimensions – is primarily a normative attempt and test constructors must try to develop procedures which satisfy strict, logically necessary, and sufficient measurement structures.

The linear logistic model (Scheiblechner, 1972; Fischer, 1973) is a true Rasch model; it presupposes the validity of the dichotomous model and postulates an additional regression on the item parameters. It is very useful for the construction of item universes or item pools (Hornke & Habon, 1986; Kubinger, 2003; Wilson & de Boeck, 2004; Kubinger, 2009).

To summarize, if the data are dichotomous, the dichotomous Rasch model is almost inevitable for test construction. But do the data have to be dichotomous?

2.1 Polytomous rating data

We now assume that the subject/observer is able to give graded responses on an ordinal scale (rating scale). There are meaningful polytomous graded response models such as the graded response model (GRM) by Samejima (1969) and the sequential model (SM) by Tutz (1990).

Let $\psi(x)$ be the logistic function

$$\psi(x) = \frac{\exp(x)}{1+\exp(x)} .$$

GRM:

$$p\left(x_{ij}=1\middle|\xi\right)=\begin{cases}1-\psi(\alpha_i\xi-\beta_{i1}) & j=0\\ \psi\left(\alpha_i\xi-\beta_{ij}\right)-\psi\left(\alpha_i\xi-\beta_{i(j+1)}\right) & 0<j<m\\ \psi\left(\alpha_i\xi-\beta_{im}\right) & j=m\end{cases}$$

SM:

$$p\left(x_{ij}=1\middle|\xi\right)=\begin{cases}1-\psi(\alpha_i\xi-\beta_{i1}) & j=0\\ \prod_{h=1}^{j}\psi\left(\alpha_i\xi-\beta_{ih}\right)\left[1-\psi\left(\alpha_i\xi-\beta_{i(j+1)}\right)\right] & 0<j<m\\ \prod_{h=1}^{m}\psi\left(\alpha_i\xi-\beta_{ih}\right) & j=m\end{cases}$$

These models are meaningful measurement models; the response variable j is used as an ordinal variable (unique except for positive monotone transformations). They are no Rasch models in the present sense, because they are not specifically objective, but they are applied unjustifiably rarely in view of the alternative models.

The generalized partial credit model (GPCM) by Muraki (1992), with its special cases of the partial credit model (PCM) by Masters (1982) and rating scale model (RSM) by Andrich (1978) are as follows

GPCM (Muraki, 1992):

$$p\left(x_{ij}=1\middle|\xi\right)=\frac{\exp(j\alpha_i\xi-\beta_{ij})}{1+\sum_{h=1}^{m}\exp(h\alpha_i\xi-\beta_{ih})}$$

(according Masters' model: $\alpha_i = 1$; and according to Andrich's model $\beta_{ij} = \beta_i$) is not a meaningful measurement model. If the response variable j is an ordinal variable or nominal, as it should be for measurement, then the addition of a constant $c > 0$ (an admissible transformation for an ordinal variable) adds a term $c\alpha_i\xi$ in the exponent which cannot be compensated by an additive constant of β independently of ξ. Therefore the probability of j changes. A model which is not invariant under admissible transformations cannot be a meaningful measurement model. If j is a rational scale (unique except for multiplication with a constant $c > 0$), then the expression is meaningful and might correctly describe the empirical distribution of some variable which can be measured on at least a rational scale level to begin with (in which case we need no measurement model).

To sum up, there are no meaningful Rasch models for rating scale data (see also below), but there are models which are used nonsensically (GPCM) and there are other models which are unjustifiably neglected (GRM, SM).

2.2 Unidimensionality

The assumption is that the complete parameter space is unidimensional.

The parameter space ξ (intelligence, for example) is *complete* if the distribution of x given ξ and arbitrary further variables y (sex, for example) is identical to the distribution of x given ξ alone (then y does not contribute to the knowledge of x given ξ, and ξ is sufficient):

$$f(x|\xi, y) = f(x|\xi) \text{ for arbitrary } y$$

(knowledge of the word 'hangar', given vocabulary of subjects, is not independent of sex; the vocabulary test is not unidimensional, because for example boys more often know 'hangar' and girls more often know 'lily').

Let us drop the assumption of unidimensionality for a moment and consider the polytomous (multicategorical) multidimensional Rasch model with the response function:

$$p\left(x_{vi}^{(h)} \middle| \xi_v^{(h)}, \sigma_i^{(h)}\right) = \frac{\exp\left[x_{vi}^{(h)}(\xi_v^{(h)} - \sigma_i^{(h)})\right]}{\sum_l \exp(\xi_v^{(l)} - \sigma_i^{(l)})} \qquad x_{vi}^{(l)} \text{ selection vector}$$

with minimal sufficient statistics

$$\sum_i x_{vi}^{(h)} \quad \text{for all } v, h, \qquad \sum_v x_{vi}^{(h)} \quad \text{for all } i, h.$$

This is a specifically objective model which is used much too rarely in practice.

If the parameters ξ and σ are linear functions of unidimensional parameters, we get the unidimensional special case with response function

$$\frac{\exp\left[\sum_l x_{vi}^{(l)}(\varphi^{(l)}\xi_v + \phi^{(l)} - \psi^{(l)}\sigma_i)\right]}{\sum_l \exp(\varphi^{(l)}\xi_v + \phi^{(l)} - \psi^{(l)}\sigma_i)}$$

$\varphi^{(l)}, \psi^{(l)}$ scoring functions,

$\phi^{(l)}$ structure function

and sufficient statistics :

$\sum_i \sum_l x_{vi}^{(l)} \varphi^{(l)}$ for all v, $\sum_v \sum_l x_{vi}^{(l)} \psi^{(l)}$ for all i

$\sum_v \sum_i x_{vi}^{(l)}$ for all l.

According to Rasch, the scoring and the structure functions cannot be estimated specifically objectively, and therefore this is not a Rasch model in the present sense. We may not simply set $\varphi^{(l)} = \psi^{(l)} = l$ and $\phi^{(l)} = 0$, and may not consider the sufficient statistic equal to the raw score (and thus get a "rating scale model"), because these values just constitute the measurement problem.

Remark: I doubt whether there are true multimensional measurements in physics which are more than a collection of isotropic measurements (e.g. the three space coordinates). However, the perceived colour space may be an intrinsic multidimensional space.

2.3 Continuity

Rasch assumes that the item characteristic of the dichotomous model has the properties of a strictly increasing continuous c.d.f.

$$p(x|\xi + \varepsilon) > p(x|\xi) \quad \text{if} \quad \varepsilon > 0$$

$$p(x|\xi) \geq 0$$

$$\lim_{\xi \to -\infty} p(x|\xi) = 0 \qquad \lim_{\xi \to \infty} p(x|\xi) = 1$$

Then he assumes that the item characteristic is twice continuously differentiable or smooth.

Remark: I doubt that there are continuous biological or psychological features (e.g. for the seemingly continuous skin complexion there are a finite number of alleles and discrete genes, and 'intelligence' may be the availability of discrete bits of information).

2.4 Local stochastic independence

The responses are independent given parameters:

$$p(\mathsf{x}|\mathsf{v},\mathsf{T}) = \prod_i p(x_{vi}|v,i) \qquad \mathsf{x} \ldots \textit{response vector}, \qquad \mathrm{T} \ldots \{I_1, I_2, \ldots, I_k\} \text{ Test.}$$

The d-aspect or d-dimensional ISOP models (presented below) allow for sets of $d \le 3$ locally dependent items.

Jannarone (1986) allows for configural scoring (e.g. the response vectors [1,1,1], [1,0,2], [0,0,3], … may have different meaning in spite of the same raw score).

The dynamic model of Kempf (1972) allows for temporal dependencies on the answers of the preceding response vector:

$p(x_{vi} = 1 | x_{v1}, x_{v2}, \ldots x_{vi-1})$ logistic function of subject and item parameters and of previous answers.

They are presumably applicable predominantly in uniformly repeated experimental situations. Their use in differential psychology models for "testlets", as used for reading comprehension, would be of interest.

2.5 Specific objectivity

Andersen (1973) and Fischer (1974) restricted specific objectivity to models with (minimal) sufficient statistics and conditional inference. Irtel (1987) suggested ordinal specific objectivity. The core of specific objectivity is that the expectations of the estimates of the subject parameters do not depend on the sample of items involved in the estimation and vice versa (the variance of the estimates may depend on the sample). The formal technique (sufficient statistics, conditional inference) of how this sample independence or to say "sample freeness" is achieved is considered less important in the following.

Definition: An instrumental paired comparison system is called '*sample free*' or '*sample independent*' if the expectations of comparisons on $A \times A$ do not depend on the elements selected from Q and the expectations of comparisons on $Q \times Q$ do not depend on the elements selected from A.

The mixed Rasch model of Rost (1996) is not a Rasch model in the present sense, because the existence of several subpopulations (or classes) of subjects with distinct item parameters is in diametric opposition to specific objectivity.

$$p(x_{vi} = 1 | v \in g) = \frac{\exp(\xi_{vg} - \sigma_{ig})}{1 + \exp(\xi_{vg} - \sigma_{ig})} \qquad g \ldots\ldots \text{classes}$$

His computer program MIRA can still be used, first in order to statistically test whether a two-or-more-class model fits the data better than the Rasch model (model control) and second in order to perhaps find two or more manifest subpopulations where different Rasch models apply (heuristic definition of populations).

3. Advantages of Rasch models for theory and practice

Traditional measures of validity and reliability are sample-dependent and not appropriate concepts to describe the quality of tests constructed by Rasch models. These tests are neither more nor less reliable or valid than traditional tests. The concept of *statistical information* outperforms by far the concept of these measures. The concept of a *measurement structure* makes the fundamental problems of psychological measurement much clearer than "(test theoretic) validity": What is a subject population? What is an item universe? What is a latent dimension?

An example of an item universe is the set of logical assertions that can be formed by "and," "or," "not," "implies" and the like, and can be translated into a non-technical form suited for psychological presentation (e.g. a graphical form, Scheiblechner, 1972). A subject population is all beings (human or not) that use assertions for communication. A latent ability is the ability to correctly use this aspect of logic. All beings using assertions must necessarily use this logic. Analphabetic nomadic children in Afghanistan as well as school children in Afghanistan and in Germany could be tested using the same Rasch-model-fitting matrices test (cf. Stori, 1985). Turkish school children in Turkey or in Germany and German school children could be measured on the same scale (cf. Sümbül, 1978). The concept of sample independence allows for the assessment of much broader subject populations and item universes than the classical sample-dependent test criteria. This gain in the precision of concepts of latent dimensions and of generalizeability of psychological assessment procedures is the true achievement of Rasch models.

4. Nonparametric models (ISOP)

I omit Rasch models for quantitative variables like the Poisson model for frequencies and exponential models for continuous variables and instead drop the assumptions of dichotomous indicators, of continuity, and of exponential families and retain unidimensionality, isotonicity (monotonicity) and local independence to obtain the isotonic psychometric models (ISOP-models, Scheiblechner, 1995, 1999, 2007; Irtel & Schmalhofer, 1982).

The axioms of the ordinal ISOP-models are:

Definition. A probabilistic paired comparison system $<A \times Q, \mathcal{F}>$, where $\mathcal{F}$ is a family of *d*-dimensional c.d.f.s indexed by $(A \times Q)$, is *weakly instrumentally independently ordered,* or a d-ISOP, if and only if the following axioms are satisfied:

W1. if $F_{vi}(.) \prec F_{wi}(.)$ for some $i \in Q$, then
$F_{vj}(.) \preceq F_{wj}(.)$ for all $j \in Q$

(*weak subject independence*).

W2. if $F_{vi}(.) \prec F_{vj}(.)$ for some $v \in A$, then

$F_{wi}(.) \preceq F_{wj}(.)$ for all $w \in A$

(*weak item independence*).

LI. The joint c.d.f. of responses of subject v to items $i,j,k,...$ is given by:

$F_v(\mathbf{x},\mathbf{y},\mathbf{z},\ldots|i,j,k,...) = F_{vi}(\mathbf{x}) \cdot F_{vj}(\mathbf{y}) \cdot F_{vk}(\mathbf{z})\ldots$

(lo*cal independence*).

(where $F(.) \prec G(.)$ and $F(.) \preceq G(.)$ is strict and weak stochastic dominance of distributions)

Theorem 1. A finite system $< A \times Q, \mathcal{F}>$ satisfies the axioms of a weakly instrumentally independently ordered system if and only if there exist real functions φ_A, φ_Q such that for all $v,w \in A$ and all $i,j \in Q$

a. $\varphi_A(v) < \varphi_A(w) \Leftrightarrow F_{vi}(.) \prec F_{wi}(.)$ for some i and

b. $\varphi_Q(i) < \varphi_Q(j) \Leftrightarrow F_{vi}(.) \prec F_{vj}(.)$ for some v.

φ_A, φ_Q are unique up to monotone increasing transformations.

The scale φ_A is usually called subject parameter and often denoted by θ or ξ and the negative scale φ_Q is called item parameter and often denoted by δ or σ in IRT. The scales are ordinal scales.

A d-ISOP or d-aspect model allows for up to $d \le 3$ ordinal reactions or ratings per item, e.g. the speed and the correctness of a reaction. Adjacent items may be locally dependent on each other. The optimal scoring function for a d-ISOP is no longer the raw score and the sum of raw scores but the *(modified) percentile scores*. They are defined for single items or for d-dimensional response vectors as

$$\text{(modified) percentile scores} = \frac{n_+ - n_-}{n_+ + n_-}$$

where n_+ is the number of responses or response vectors which are inferior to (smaller than) the given response or vector and n_- is the number of superior responses. The modified percentile scores are maximum likelihood estimates of the ordinal positions of subjects or items and their order is sample independent.

If interval scales for subjects and items are desired, then a further axiom (Co, *cancellation of order o*) is needed.

Definition. A d-ISOP is a d-ADISOP, a *d component additive conjoint, isotonic, probabilistic model*, if and only if in addition to the axioms of d-ISOP the following axiom is satisfied (Scheiblechner, 1999):

(Co): *validity of the additively implied inequalities on the c.d.f.s indexed by* $A \times Q$

The order relations on the c.d.f.s, indexed by $A \times Q$ implied by an additive representation of the response variables in all sets of 2 to o = $\min\{n,k\}-1$ stochastic dominance relations (antecedents), must be valid.

(cancellation of order o = $\min\{n,k\}-1$)

If an interval scale for the rating response is desired, then the cancellation of the subject parameter and the rating scale and the cancellation of the item parameter and the rating scale is additionally needed.

Definition. A d-ADISOP is a complete d-ADISOP or d-CADISOP if and only if (Co) is dropped in favour of (W4) and (RS):

(W4): the probabilities of the d-dimensional c.d.f.s indexed by $A \times Q$ are isotonically ordered:

if $F_{vi}(x_1, x_2, \ldots, x_d) \le F_{vi}(x_1', x_2', \ldots, x_d')$ for some v and some i,

then $F_{wj}(x_1, x_2, \ldots, x_d) \le F_{wj}(x_1', x_2', \ldots, x_d')$ for every w and j.

(weak instrumental variable independence)

(RS): (restricted solvability)

The usual scoring function by the raw score presupposes a very implausible special case of a d-CADISOP where all measurement units (across ratings, across items, across subjects resulting in the same raw score – e.g. the response vectors (1,1,1), (1,0,2), (0,0,3) - defined to correspond to the same interval scale value of the latent dimension) are assumed to be equal in addition to the above axioms.

The ISOP-models were applied to the MR SOC test (sense of coherence; Scheiblechner & Lutz, 2009; n = 1156 subjects, 10 items of positive and 10 items of negative feelings). The weak independence axioms W1 and W2 could not be rejected (by generalized isotonic regressions). The cancellation of order o, Co, of subjects and items is rejected by a heuristic likelihood ratio test but seems acceptable by a Schwarz-Bayes information criterion for model fit. The weak instrumental variable independence W4 is definitely rejected. The rating scale of the response (1 never, 2 rarely, 3 rather frequently, 4 frequently) interacts with the subject parameter; emotionally stable subjects use the frequency terms of emotional states differently than insecure subjects. The rating scale of the response also interacts with the item content. The frequencies of strong emotions are judged differently than the frequencies of more commonplace feelings. The usual raw score or simple sum score of test scoring is not acceptable. The modified percentile scores, however, give sample independent estimates of ordinal positions.

This increasing degradation of models can be grasped by a sequence of conditional graphical model tests, where the fit of a higher model (an additional axiom) is plotted against the preceding weaker model. If the additional axiom is valid, the graph must be smooth and isotonically increasing. The ordinates of the graphs are estimated relative frequencies. The deviations of the W1W2Test in Figure 1 (model ISOP) are small compared to the random fluctuations of relative frequencies. The deviations of the CoTEST are more pronounced,

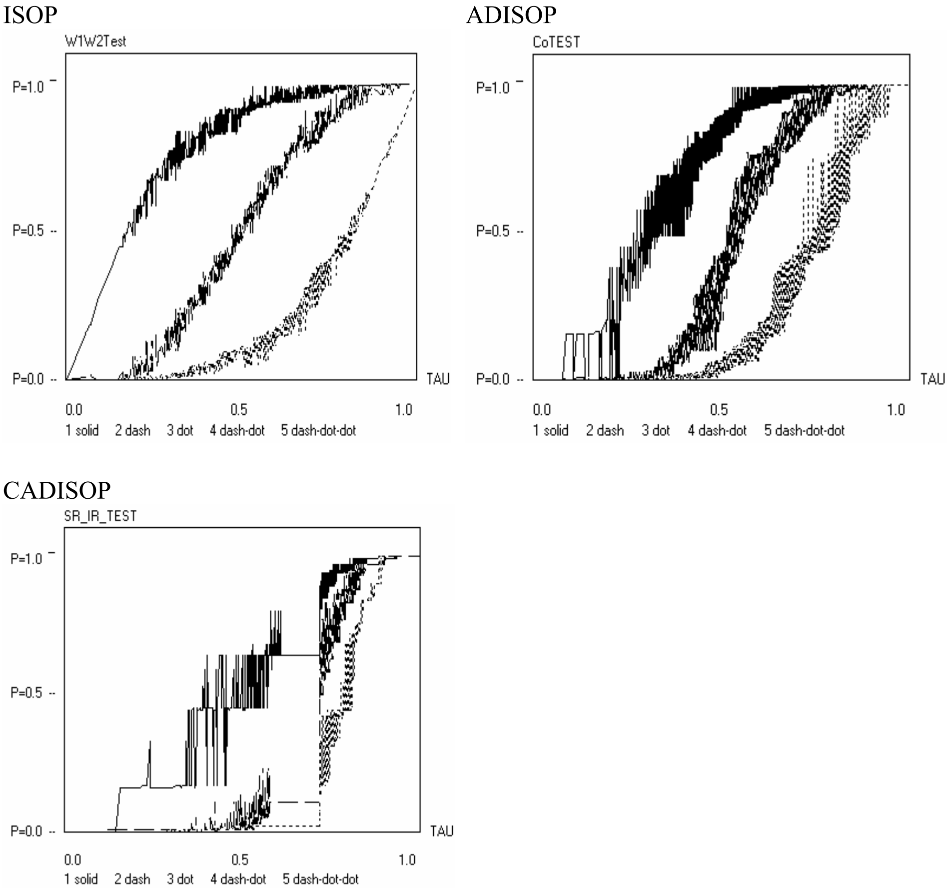

Figure 1:

Graphical model controls. Tests of axioms conditional on the validity of the subordinate model. The ordinates are estimated relative frequencies showing the easiness of passing from a lower response category to the next. If the axioms are perfectly valid, all graphs should be smooth isotonically increasing functions. The ISOP graph tests W1 and W2. The ADISOP graph tests Co conditional on ISOP. The CADISOP graph tests W4 conditional on ADISOP

especially at the left end (low) of the never/rarely step function and the right end (high) of the rather frequently/frequently step function (model ADISOP). The graph of the SR_IR_TEST has a disturbing gap at approximately 0.6 and a wide vertical spread of the 3 step functions. The model CADISOP is definitely rejected. This rejects all parametric IRT models with additive subject and item parameters and the simple sum score of test evaluation.

Figure 2 shows the item step response functions ISRF (threshold functions for passing frequency response categories) and the category characteristic curves CCC (probabilities of response categories) of the ISOP model for the total MR-SOC test of 20 items. They look similar to the corresponding curves of parametric models.

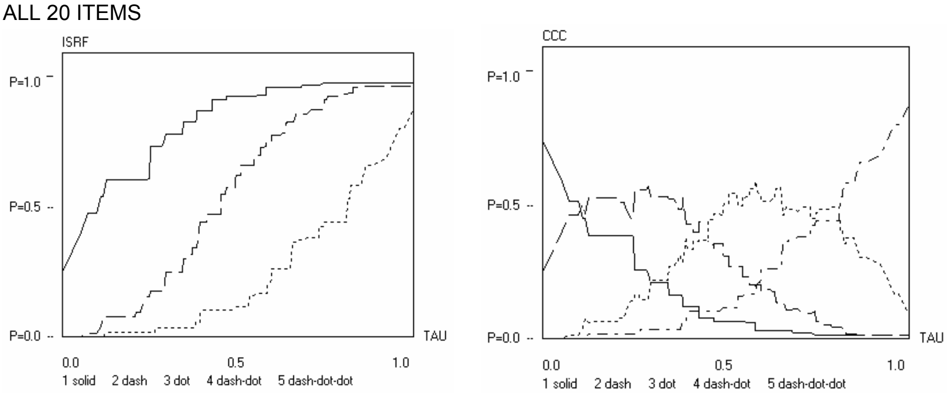

Figure 2:

ISRFs, item step response functions (threshold functions between adjacent categories) and CCCs, category characteristic curves (probabilities of response categories as functions of the ordinal parameter of the subject-item pair) for the ISOP model

5. Summary

The following models have been discussed:

The dichotomous logistic model of Rasch:
the ideal prototype of a fully specifically objective model, measurement independent of instruments used (sample independent)

2PL model:
with item discrimination parameters; not identified (more free parameters than degrees of freedom in the data)

3 PL and extensions:
additional guessing parameter and further parameters; see Birnbaum

LLTM:
dichotomous Rasch model plus explanatory linear model for item parameters;

GRM, SM:
meaningful polytomous measurement models (rating scale as ordinal scale), not specifically objective

GPCM, PCM, RSM:
not meaningful polytomous (rating scale) measurement models, distribution models for rating scale data

Polytomous Rasch model:
nominal model, multidimensional, fully specifically objective; unidimensional special cases not specifically objective

Kempf, Jannarone dynamic models:
(restricted) local dependencies admissible, experimental settings

Mixed Rasch model:
not a Rasch model, heuristic tool

ISOP:
non parametric, sample independent; ordinal, interval extensions

Rasch has stimulated the production of many models which do not always conform to his postulates and whose applications are not by far fully explored.

References

Andersen, E. B. (1973). *Conditional inference and models for measuring*. Copenhagen: Mentalhygiejnisk Forlag.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta psychologica, 37*, 359-374.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests [Introduction into the theory of psychological tests]*. Bern: Huber.

Hornke, L. F., & Habon, M. W. (1986). Rule-based item bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement, 10*, 369-380.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika, 51*, 357-373.

Kempf, W. F. (1972). Probabilistische Modelle experimentalpsychologischer Versuchssituationen [Probabilistic models of experimental psychological situations]. *Psychologische Beiträge, 14*, 16-37.

Kubinger, K. D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie [Critical evaluation of latent trait theory]. In K.D. Kubinger (Ed.), *Moderne Testtheorie – Ein Abriß samt neuesten Beiträgen* [Modern psychometrics – A brief survey with recent contributions] (pp. 19-83). Munich: PVU.

Kubinger, K. D. (2003). Adaptives Testen [Adaptive testing]. In K.D. Kubinger and R.S. Jäger (Eds.), *Stichwörter der Psychologischen Diagnostik* [Key words of Psycho-diagnostics] (pp. 1-9). Weinheim: PVU.

Kubinger, K. D. (2009). Applications of the Linear Logistic Test Model in Psychometric Research. *Educational and Psychological Measurement, 69,* 232-244.

Irtel, H. (1987). On specific objectivity as a concept in measurement. In E. E. Roskam & R. Suck (Eds.), *Progress in Mathematical Psychology* (Vol. 1, pp. 35-45). Amsterdam North-Holland: Elsevier.

Irtel, H., & Schmalhofer, F. (1982). Psychodiagnostik auf Ordinalskalenniveau: Messtheoretische Grundlagen, Modelltest und Parameterschätzung [Psychodiagnostics on ordinal scale level: Measurement theoretic foundations, model test and parameter estimation]. *Archiv für Psychologie, 134*, 197-218.

Lutz, R., Herbst, M., Iffland, P., & Schneider, J. (1998). Möglichkeiten der Operationalisierung des Kohärenzgefühls von Antonovsky und deren theoretische Implikationen [Possible operationalizations of the sense of coherence of Antonovsky and theoretical implications]. In J. Margraf, J. Siegrist & S. Neumer (Hrsg.), *Gesundheits- oder Krankheitstheorie? Saluto- versus pathogenetische Ansätze im Gesundheitswesen* (S. 171-185). Berlin: Springer.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Muraki, E. (1992). A generalized partial credit model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*, 159-176.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceeds of the Forth Berkley Symposium on Mathematical Statistics and Probability*. 5. Berkley: University of California Press, 321-333.

Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion [Text book test theory, test construction]*. Göttingen: Huber.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika,* Special monograph, Monograph Supplement No. 17 .

Scheiblechner, H. (1972). Das Lernen und Lösen komplexer Denkaufgaben [Learning and solving complex thinking tasks]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 19*, 476-506.

Scheiblechner, H. (2007). A unified, nonparametric IRT measurement model for d-dimensional psychological test data (d-ISOP). *Psychometrika*, *72,* 43-67.

Scheiblechner, H., & Lutz, R. (2009). Die Konstruktion eines optimalen eindimensionalen Tests mittels nichtparametrischer Testtheorie (NIRT) am Beispiel des MR SOC [The construction of an optimal unidimensional test by means of nonparametric test theory (NIRT) at the example of the MR SOC}. *Diagnostica*, *55*, 41-54.

Stori, K. (1985). Entwicklung eines kulturfairen Intelligenz-Tests. [The construction of a culture fair intelligence test]. Dissertation (KT) Philipps-Universität Marburg.

Sümbül, O. (1978). Die Entwicklung der logischen Intelligenz. [The development of logical intelligence]. Dissertation (KT) Philipps-Universität Marburg.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology 43*, 39-55.

van der Linden, W. J., & Hambleton, R. K. (eds.) (1997). *Handbook of modern item response theory*. New York: Springer.

Verhelst, N. D., & Glas, A. W. (1995). The one parameter logistic model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models, foundations, recent developments, and applications.* New York: Springer.

Wilson, M., & de Boeck, P. (2004). Descriptive and explanatory item response models. In P. de Boeck & M. Wilson (eds.), *Explanatory item response models* (pp. 43-74). New York: Springer.