Mathematics and Statistics 10(3): 690-700, 2022 http://www.hrpub.org
DOI: 10.13189/ms.2022.100326

Form Invariance — An Alternative Answer to the Measurement Problem of Item Response Theory

Henrik Bernshausen[1], **Christoph Fuhrmann**[2], **Hanns-Ludwig Harney**[3], **Klaus Harney**[4], **Andreas Müller**[5,*],

[1]Faculty of Science and Technology, Physics Education Group, University of Siegen, D-57068 Siegen, Germany

[2]School of Education, Institute for educational research, University of Wuppertal, D-42097 Wuppertal, Germany

[3]Max-Planck-Institute for Nuclear Physics, Postfach 103980, D-69029 Heidelberg, Germany

[4]Faculty of Philosophy and Educational Science, Institute for educational research, Ruhr-Universit¨at, D-44801 Bochum, Germany

[5]Faculty of Sciences/Physics Department and Institut Universitaire de Formation des Enseignants, University of Geneva, CH - 1211 Geneva 4, Switzerland

Received January 25, 2022; Revised March 25, 2022; Accepted April 15, 2022

Cite This Paper in the following Citation Styles

***(a)**: [1] Henrik Bernshausen, Christoph Fuhrmann, Hanns-Ludwig Harney, Klaus Harney,Andreas Müller, "Form Invariance — An Alternative Answer to the Measurement Problem of Item Response Theory," Mathematics and Statistics, Vol.10, No.3, pp. 690-700, 2022. DOI: 10.13189/ms.2022.100326*

***(b)**: JHenrik Bernshausen, Christoph Fuhrmann, Hanns-Ludwig Harney, Klaus Harney,Andreas Müller, (2022). Form Invariance — An Alternative Answer to the Measurement Problem of Item Response Theory. Mathematics and Statistics, 10(3), 690-700. DOI: 10.13189/ms.2022.100326*

Abstract The measurement problem of item response theory is the question of how to assign ability parameters $\boldsymbol{\theta}$ to persons and difficulty parameters $\boldsymbol{\sigma}$ to items such that the comparison of abilities is independent of the specific set of difficulties $\boldsymbol{\sigma}$. Correspondingly, the comparison of difficulties $\boldsymbol{\sigma}$ should be independent of the specific set of abilities $\boldsymbol{\theta}$. These requirements are called specific objectivity. They are the basis of the Rasch model. It measures $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ on one and the same scale.

The present paper asks the different question of how to assign ability parameters $\boldsymbol{\theta}$ to persons in a way that the comparison of abilities is independent of the position on the scale where the measurement takes place. Correspondingly, the comparison of difficulties $\boldsymbol{\sigma}$ should also be independent of the position on the scale where the calibration of difficulties takes place. Again, $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ measured on one and the same scale. These requirements are called form invariance. They lead to an item response function (IRF) different from that of the Rasch model. It integrates information from θ and σ beyond the mere score dependence and also shows specific objectivity (in a generalized mathematical form). The properties of the form invariant item response function are compared to that of the Rasch model, and related to previous work by Warm, Jaynes and Samejima. Moreover, several numerical examples for the use of it are provided.

Keywords Item Response Theory, Form Invariance, Invariant Measure, Individual Answer Patterns

1 Introduction

The "measurement problem" in the title of the present contribution refers to a question raised by G. Rasch when laying the foundations to a model that nowadays carries his name [29, 30, 34, 35], a model of psychological measurement developed to a high degree of methodological sophistication [21], with very broad applications e.g. in international large scale studies [1]; science education [22, 28]; and the educational [40] and human sciences [7] in general. He asked the question: Is it possible to assign ability parameters $\boldsymbol{\theta}$ to persons and difficulty parameters $\boldsymbol{\sigma}$ to items such that the comparison of abilities does not depend on the items used in the test — and that the comparison of items does not depend on the ability parameters? Rasch used the term "specific objectivity" for this requirement, see ref. [31, 33, 34, 35] as well as further investigations by G.H. Fischer [11, 12, 13]. For some recent accounts and applications, see [9, 19]. Specific objectivity shall make sure that the results of measurements are independent of the instrument used for a measurement. Rasch created specific objectivity via the definition of a scale common to all parameters of item response theory (IRT). In his work the logistic item response function (IRF) sets this common scale. We call this the logistic or Rasch model.

The present approach starts out from asking a question different form the one leading to the Rasch model: Is it possible to assign ability parameters θ to persons in a way that the comparison of abilities is independent of the position on the scale where the measurement takes place (and analogously for the comparison of difficulties σ)? In other context of measurement this appears as a very plausible requirement, lengths (differences of positions) are independent of the position on the yardstick, energy differences are independent of the location on the energy scale, and so on. This basic requirement about measurement is put into a formal framework which in mathematics and science is widely considered appropriate to treat invariance properties. This is group theory. The present work is based on the group theoretical concept of form invariance, a general approach to models of probabilistic inference, see Chap. 12 of Ref. [16].

To require form invariance suffices to define the IRF. This occurs because group theory introduces a measure, i.e. the inverse unit of length, in the space of the parameter(s). Since parameters shall be compared by taking their difference the measure must be constant. From this requirement, it is inferred that the IRF should be a squared sine-function instead of the logistic function of the Rasch model, hence the name "trigonometric model".

The paper is organized as follows. In Sec. 2 the dichotomous IRT is briefly reviewed. This provides the subsequent notation. In Sec. 3 the Rasch model is described. We summarize the argument of specific objectivity and state the model explicitly. Sec. 4 introduces the trigonometric model. In Sec. 4.1 we define form invariance as well as the invariant measure in the parameter space. Sec. 4.2 states the model explicitly together with the appropriate system of ML equations. In Sec. 5 data generated by Monte Carlo simulation are interpreted. We generate data by use of the logistic model and subsequently analyze them by both, the logistic and the trigonometric models. We show that the latter one makes use of the individual patterns of answers. Section 6 summarizes and concludes the paper. A glossary of the notations and abbreviations can be found in Appendix A.

2 Preliminaries: The Structure of Dichotomous Item Response Functions

The dichotomous item response model is defined by the product

$$q^{\text{tot}}(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{\sigma}) = \prod_{p=1}^{N_P}\prod_{i=1}^{N_I} q(x_{p,i}|\theta_p,\sigma_i) \tag{1}$$

of the binomial models

$$q(x_{p,i}|\theta_p,\sigma_i) = [R(\theta_p,\sigma_i)]^{x_{p,i}}\,[1-R(\theta_p,\sigma_i)]^{1-x_{p,i}}\ . \tag{2}$$

Here, the function $R(\theta_p,\sigma_i)$ is called item response function (IRF). The symbol q means the probability to obtain the answer $x_{p,i}$ of the p-th person to the i-th item. There are two possible answers $x_{p,i}=0$ or $x_{p,i}=1$ which can be interpreted as the alternative of false and true. The fact that q^{tot} factorizes into these probabilities expresses the common assumption that $x_{p,i}$ is statistically independent of $x_{p',i'}$ for $(p,i)\neq(p',i')$. We use the notation

$$\boldsymbol{\theta} = \begin{pmatrix}\theta_1\\ \vdots\\ \theta_{N_P}\end{pmatrix}, \qquad \boldsymbol{\sigma} = \begin{pmatrix}\sigma_1\\ \vdots\\ \sigma_{N_I}\end{pmatrix} \tag{3}$$

for the vectors of the person and item parameters as well as

$$\boldsymbol{x} = (x_{p,i})\,; \qquad p=1,\dots,N_P; \qquad i=1,\dots,N_I \tag{4}$$

for the matrix of the answers $x_{p,i}$. Given R, the model (1) depends parametrically on the person parameters $\boldsymbol{\theta}$ and the item parameters $\boldsymbol{\sigma}$. This allows to infer the parameters from the data matrix $\boldsymbol{x}$.

3 The Rasch Model

The Rasch model has been constructed to convey specific objectivity to the person parameters $\boldsymbol{\theta}$ as well as to the item parameters $\boldsymbol{\sigma}$. This means that the comparison between σ_i and $\sigma_{i'}$ should not depend on $\boldsymbol{\theta}$ and the comparison between θ_p and $\theta_{p'}$ should not depend on $\boldsymbol{\sigma}$. This "specific objectivity" has been postulated by Rasch [31] and investigated by G.H. Fischer, see [11, 12] and p. 527 of [13]. Ability and difficulty are conjointly measured on the same scale. From this, it has been concluded that R must be a function of the differences $\theta_p-\sigma_i$ of ability and difficulty parameters, see [5, 35] and theorem 5 on p. 529 of [13]. In order to guarantee that the probability of solution increases with increasing ability it seemed natural to require R to be a monotonic function of $\theta_p-\sigma_i$.

Specific objectivity does not determine the IRF uniquely. For this an additional requirement must be introduced such as the scores to be the sufficient statistics of the model, see [12] and p. 526 of [13]. The scores are the numbers of correct answers given by a person or given to an item. Rasch [32, 35] arrived at the so-called logistic IRF

$$R^{\text{lg}}(\xi) = \frac{\exp(\xi)}{1+\exp(\xi)}\,, \tag{5}$$

where $\xi=\theta_p-\sigma_i$ may assume any real value. This function defines the Rasch model. Given Eq. (5) and $\xi=\theta_p-\sigma_i$ we obtain the explicit form of the model from Eqs. (1,2) ,

$$\begin{aligned} q^{\text{lg}}(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{\sigma}) &= \prod_{p=1}^{N_P}\prod_{i=1}^{N_I}\left[R^{\text{lg}}(\theta_p-\sigma_i)\right]^{x_{p,i}} \\ &\qquad \left[1-R^{\text{lg}}(\theta_p-\sigma_i)\right]^{1-x_{p,i}} \\ &= \prod_{p=1}^{N_P}\prod_{i=1}^{N_I}\Big[\exp\Big(x_{p,i}(\theta_p-\sigma_i)\Big)\Big] \\ &\qquad [1+\exp(\theta_p-\sigma_i)]^{-1} \\ &= \exp\left(\sum_{p=1}^{N_P} t_p\theta_p - \sum_{i=1}^{N_I} s_i\sigma_i\right) \\ &\qquad \prod_{p,i}\left(1+\exp(\theta_p-\sigma_i)\right)^{-1}. \end{aligned} \tag{6}$$

The third version of this equation shows that $\boldsymbol{x}$ enters into the Rasch model via the person scores

$$t_p = \sum_{i=1}^{N_I} x_{p,i} \tag{7}$$

and the item scores

$$s_i = \sum_{p=1}^{N_P} x_{p,i}\,. \tag{8}$$

The inferred parameters will depend exclusively on the scores. In other words, the scores provide a set of sufficient statistics of the model.

Since the model depends on the differences $\theta_p - \sigma_i$, it remains unchanged if one adds the same constant to each of its parameters. This freedom of the zero point of the scale is removed by a convention such as assigning the value of zero to one of the parameters. We choose

$$\sum_{i=1}^{N_I} \sigma_i = 0\,. \tag{9}$$

This convention is incorporated, e.g., in the computer code ConQuest [2] used below in Sec. 5.

In general, statistical parameters are estimated by the method of maximum likelihood (ML), i.e. as the point which maximizes the probability of the observed data. This method is due to R.A. Fisher [14]. By asking the logarithmic derivative of q^{lg} to vanish, one finds for the Rasch model the ML equations

$$\begin{aligned} t_p &= \sum_i R^{\mathrm{lg}}(\theta_p - \sigma_i)\,, \quad p = 1, \ldots, N_P \\ s_i &= \sum_p R^{\mathrm{lg}}(\theta_p - \sigma_i)\,, \quad i = 1, \ldots, N_I\,. \end{aligned} \tag{10}$$

Hence, the ML estimators of the logistic model are determined by the scores of the persons and items. The first of Eqs. (10) entails by the monotony of R^{lg} that two persons p, p' having the same score $t_p = t_{p'}$ have the same ability estimator $\theta_p = \theta_{p'}$. The corresponding holds for the difficulty parameters. This does, however, not mean that the person parameter θ_p would depend on t_p alone and decouple from the item parameters. One sees that a person parameter still depends on all item parameters; an item parameter depends on all person parameters.

In practice, the Eqs. (10) are not always used. The results may be unstable. The reason is, according to Rasch [30], Andersen [3, 4] and Fischer (see p. 260 of [10] and p.534 of [13]), that the ML estimators do not converge towards well defined values when, say, N_P grows indefinitely while the items stay unchanged. This may not be a well defined limit because the number of parameters increases with the number of persons, i.e. the model changes. The additional person parameters may counteract the gain of precision expected for the item parameters. In this case the ML estimators are said to lack consistency [26]. This lack is ascribed to the "incidental" parameters (the person parameters in the above) when the interest focusses on the "structural" parameters (the item parameters).

We now describe a model of IRT which fits into general statistics. Its full set of estimators can be found by usual ML estimation.

4 The Trigonometric Model: Main Results

In the present section the concept of form invariance is introduced (in Sec. 4.1); the trigonometric model follows from it (in Sec. 4.2).

4.1 Form Invariance

Consider the binomial model of Sec. 2,

$$q(x|\theta,\sigma) = [R(\theta,\sigma)]^x\,[1 - R(\theta,\sigma)]^{1-x}\,. \tag{11}$$

To convey the idea of form invariance we consider only a single person and a single item, characterized by the parameters θ and σ, respectively. The bold face letters (denoting vectors) in Sec. 2 are thus replaced by ordinary letters without an index. For simplicity we only consider the ability parameter to be inferred on the basis of x; the difficulty σ is taken as known and is omitted from the consideration. It is re-introduced later.

The idea of invariance of the measurement (here of ability) with respect to the location (absolute value) on the measurement scale will be expressed by group theory, a well-established mathematical language when dealing with invariance. The reasoning proceeds in two steps, the introduction of the symmetry group and the requirement of constant measure.

4.1.1 Introduction of the Invariance Group

In order to introduce the desired invariance we relate the probability distribution q to the inferred parameter θ via a mathematical group of transformations $\mathbf{G}_\theta$. If $\mathcal{G}$ is a group in the sense of group theory the distribution q presents a symmetry. This symmetry we call "form invariance". The properties of a group are mirrored by the properties of parameters, see Chap. 6.1 of [16]. This leads, in the present context, to parameters which can be added (for a shift) and subtracted (for comparison).

The group is constructed as follows. One represents the distribution $q(x|\theta)$ by a vector $\boldsymbol{a}(\theta)$ with the components $a_x(\theta)$ labelled by the possible events $x = 0, 1$. The vectors shall be related to each other by the action of an element $\mathbf{G}$ of the group; i.e. any $\boldsymbol{a}$ can be obtained from any other $\boldsymbol{a}'$ by the action of an appropriate element of $\mathcal{G}$. In particular, one can choose an "initial" vector $\boldsymbol{a}(\theta = 0)$ such that $\boldsymbol{a}(\theta)$is obtained by applying $\mathbf{G}_\theta$,

$$\boldsymbol{a}(\theta) = \mathbf{G}_\theta\,\boldsymbol{a}(0)\,. \tag{12}$$

The choice of $\boldsymbol{a}(0)$ does not entail a loss of generality since any $\boldsymbol{a}'(0)$ could be used via the action of the group element that shifts $\boldsymbol{a}(0)$ to $\boldsymbol{a}'(0)$.

The transformations $\mathbf{G}_\theta$ shall be linear. Therefore the components of the vector $\boldsymbol{a}(\theta)$ are not the probabilities $q(x|\theta)$ themselves. This is impossible for the following reason. Linear transformations of the probabilities would require matrices with positive elements only. Given a matrix $\mathbf{G}_\theta$, its inverse must be an element of the group, too, but it would not consist of positive elements only. Thus there is no group of linear transformations acting directly on the probabilities. Instead of the probabilities one rather chooses the probability- amplitudes

$$a_x(\theta) = \pm\sqrt{q(x|\theta)} \tag{13}$$

to be the elements of the $\boldsymbol{a}(\theta)$ and considers a group $\mathcal{G}$ of orthogonal transformations acting on $\boldsymbol{a}$. In the present case of two dimensions, i.e. of two possible answers to a given item and the $a_x(\theta)$ being real, there is only one group of orthogonal transformations, the rotations $\mathcal{G} = O(2)$ about the origin. Form invariance is a symmetry in the following geometrical sense: If one interprets the vector $\boldsymbol{a}(\theta)$ as comprising the Cartesian coordinates of a point in space, the set of points obtained by letting θ run over its domain of definition is a curve. Due to the properties of a group this curve remains the same under the transformations of the group $\mathcal{G}$. One sees that the binomial model possesses the symmetry of the circle. For the binomial model (11) we consider the vector

$$\boldsymbol{a}(\theta) = \begin{pmatrix} \sqrt{R(\theta)} \\ -\sqrt{1-R(\theta)} \end{pmatrix} . \tag{14}$$

Since the components of $\boldsymbol{a}$ are labelled by the values x of the possible events, their number equals the dimension of the space from which $\boldsymbol{a}$ is taken. A finite alternative thus corresponds to vectors and orthogonal transformations in a finite-dimensional space.

4.1.2 The Measure in Parameter Space

For any position θ on the above curve, one and the same change $\Delta\theta$ of the parameter should amount to the same length of the shift along the curve. Mathematically, this means that we have to choose the parameter θ such that its measure is uniform throughout the continuum of its values. This is implied by requiring an invariant comparison: All parameters of IRT should refer to the same scale, the scale of θ; and any two of them should be compared by taking their difference, independent of the location (absolute value) of the measurement scale. For this the unit of length must be independent of the values to be compared; in other words, the unit of length must be constant over the space of θ. The "measure", i.e. the inverse unit of length, is proportional to

$$\mu(\theta) = \left[\sum_x \left(\frac{\partial}{\partial\theta} a_x(\theta)\right)^2\right]^{1/2} . \tag{15}$$

This expression is known from differential geometry [20]. For the binomial model (11) it requires

$$\mu(\theta) = \left(\frac{[R'(\theta)]^2}{R(\theta)(1-R(\theta))}\right)^{1/2} . \tag{16}$$

Hence, the measure becomes independent of θ if

$$R'(\theta) \propto \sqrt{R(1-R)}\,. \tag{17}$$

Thus the requirements of (i) the existence of a symmetry group and (ii) constancy of the measure assigned to the group parameter lead to a differential equation for the IRF. It is solved by a trigonometric function squared. We set

$$R^{\mathrm{tr}}(\theta) = \sin^2(\theta + \pi/4)\,. \tag{18}$$

The shift by $\pi/4$ has been introduced to make $R^{\mathrm{tr}}(0) = R^{\mathrm{lg}}(0) = 1/2$.

Given the constant measure in the space of the parameter θ, it is meaningful to compare ability and difficulty by taking their difference such that the IRF of the two variables θ, σ becomes the function $R^{\mathrm{tr}}(\theta - \sigma)$. Then the vector $\boldsymbol{a}$ depends on $\theta - \sigma$.

4.1.3 Explicit Representation of the Symmetry Group and Inclusion of Difficulty Parameters

From the IRF (18) we explicitly obtain the transformation $\mathbf{G}_\theta$. Inserting Eq. (18) into (14) one finds that the symmetry group consists of the rotations

$$\mathbf{G}_\theta = \begin{pmatrix} \cos\theta, & -\sin\theta \\ \sin\theta, & \cos\theta \end{pmatrix} \tag{19}$$

applied to the initial vector

$$\boldsymbol{a}(0) = 2^{-1/2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} . \tag{20}$$

This reproduces the shift in Eq. (18).

We now turn to the full model including ability and difficulty parameters (θ and σ, respectively). The first step is to generalize Eqs. (14) eq. (18) to include also the difficulty parameters, i. e.

$$\boldsymbol{a}(\theta,\sigma) = \begin{pmatrix} R(\theta-\sigma) \\ -\sqrt{1-R(\theta-\sigma)} \end{pmatrix} . \tag{21}$$

with

$$R(\theta-\sigma) = \sin^2(\theta - \sigma + \pi/4) \tag{22}$$

Note that the occurrence of $\theta - \sigma$ as argument of R reflects conjoint measurement, as far as the Rasch model, see sect. 3. Form invariance is the defined analogously to Eqs. (12), (19) and (20)

$$\boldsymbol{a}(\theta,\sigma) = \mathbf{G}_{\theta,\sigma}\boldsymbol{a}(0,0) \tag{23}$$

with

$$\mathbf{G}_{\theta,\sigma} = \begin{pmatrix} \cos(\theta-\sigma), & -\sin(\theta-\sigma) \\ \sin(\theta-\sigma), & \cos(\theta-\sigma) \end{pmatrix} \tag{24}$$

and

$$\boldsymbol{a}(0,0) = 2^{-1/2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} . \tag{25}$$

Before we turn to the full multidimensional treatment in sect. 4.2 we note some properties of the trigonometric model.

4.1.4 Some Properties and Relations to other Results

(i) Specific Objectivity: We now show that form invariance implies specific objectivity. The latter is defined in a general mathematical form by [10, 34]

$$K(R(\theta,\sigma), R(\theta',\sigma)) = U(\theta,\theta') \tag{26}$$

where R is the "reaction function", an (in principle) empirically measurable function of parameters to be determined, and K is the "comparison function" (or "comparator"). Equation (26) is a functional equation, the "comparator equation". For the Rasch model, R is the item response function, and $K(x,y) = \log(x) - \log(y)$, leading to $U(\theta,\theta') = \theta - \theta'$ [10, 34] . For the form invariant model, the "reaction function" is defined by

$$R(\theta,\sigma) = \mathbf{G}_{\theta,\sigma} \tag{27}$$

and the comparator by

$$K(\mathbf{G},\mathbf{G}') = \mathbf{G}'\,\mathbf{G}^{-1}\,. \tag{28}$$

Using the group properties of G, one obtains

$$\begin{aligned} U(R(\theta',\sigma),R(\theta,\sigma)) &= \mathbf{G}_{\theta,\sigma}\,\mathbf{G}^{-1}_{\theta',\sigma} \\ &= \mathbf{G}_{\theta-\sigma}\,\mathbf{G}^{-1}_{\theta'-\sigma} \\ &= \mathbf{G}_{\theta-\sigma}\,\mathbf{G}_{\sigma-\theta'} \\ &= \mathbf{G}_{\theta-\sigma+\sigma-\theta'} \\ &= \mathbf{G}_{\theta-\theta'} \end{aligned} \tag{29}$$

not depending on σ, as required by specific objectivity. Here, the group properties for $\mathbf{G}$ were used (inverse element, line 3; group association law, line 4)

Some comments are in order:

a) Note that the reaction function (27) is also (in principle) empirically measurable, as required by the general definition of specific objectivity. In fact, $\mathbf{G}_{\theta,\sigma}$ can explicitly be written as

$$\mathbf{G} = \begin{pmatrix} R^{1/2}(\theta+\pi/4), & R^{1/2}(\theta-\pi/4) \\ R^{1/2}(\theta+\pi/4), & R^{1/2}(\theta-\pi/4) \end{pmatrix}, \tag{30}$$

thus in terms of the item response function, as in the case of the Rasch model. A proof of (30) is given in the appendix B.

b) In the Rasch model, scalar reaction functions and comparators are used, leading to the logistic IRF. The generalisation by the form invariant model is to allow for matrices and operations on them for the reaction functions and the comparator, leading to the trigonometric IRF.

c) In the form invariant model, the basic requirement is that the unit of measurement is constant across the scale, so that comparisons or differences of the parameters have the same meaning everywhere on it, specific objectivity (in the sense of a)) is then inferred. Conversely, in the Rasch model, specific objectivity is the basic requirement, and the scale level of a difference scale is a consequence, see Chap. 3.1. of [36]. Thus, both models fulfil both measurement requirements discussed above, (1) independence of the comparison of abilities from the specific set of difficulties used (and vice versa), and (2) independence of the comparison of abilities from the position on the scale where the measurement takes place (analogously for difficulties), but in a sense specific for each model. Table 1 summarizes this.

(ii) The measure μ of Eq. (15) can be expressed via the logarithm of the probability q in the two ways

$$\begin{aligned} \mu(\theta) &= \frac{1}{2}\left[\sum_x q(x|\theta)\left(\frac{\partial}{\partial\theta}\ln q(x|\theta)\right)^2\right]^{1/2} \\ &= \frac{1}{2}\left[-\sum_x q(x|\theta)\frac{\partial^2}{\partial\theta^2}\ln q(x|\theta)\right]^{1/2}. \end{aligned} \tag{31}$$

The first version follows from Eq. (15) together with the definition (13) of the components of $\boldsymbol{a}$. The second version then follows from the fact that q is normalized for all θ, see e.g. Chap. 9.1 of [16]. This measure is known as Jeffreys' rule [17, 18]. The expression in square brackets is Fisher's information [14]. Equations (31) show that constancy of the measure is equivalent to constancy of the information; i.e. the information contained in the data is independent of θ. Both, Fisher and Jeffreys, have initiated the discussion of the measure administered to statistical parameters. They have thus revived Bayesian statistics [6]. J. Rost, on p. 357 of [36], gives the second one of Eqs. (31). In the sequel he states and justifies that (within the Rasch model) the error of a parameter depends on the value of its estimator. Since the root mean square error is the inverse of μ, the measure is not uniform within the Rasch model (similarly Bühner on p. 323 of [8]).

(iii) Samejima required the Fisher information to be constant, which amounts to require constant measure. In 1980 she concluded [37] that the IRF should be a squared $\sin$-function. But she insisted on a monotonic IRF and used only the increasing part of $\sin^2$. Therefore form invariance is not implied by Ref. [37]. Samejima's approach was not taken up in practice, e. g. her article is not quoted in volume 26 of the Handbook of Statistics (see e.g. [13]) devoted to psychometrics.

(iv) Warm's correction [39] to the joint maximum likelihood estimators, designed to reduce a possible bias of them, vanishes when the Fisher information is constant. In this sense the trigonometric IRF leads to estimators with minimal bias.

(v) The expressions of Eqs. (15, 31) are not only obtained from differential geometry, they also agree with Jeffreys´ rule [17, 18] given in Chap. 9.1 of [16].

(vi) The trigonometric IRF ensures that the measure is uniform with respect to every parameter θ_p or σ_i. A result of group theory (not proven here) says that this occurs if and only if the symmetry group of the full set of parameters of the trigonometric model is Abelian. A gratifying consequence is: The trigonometric parameters are independent of each other in the sense that any parameter can be integrated out of the likelihood function without loosing information on any of the remaining ones. See below Eq. (36). To explain this last point, the explicit form of the trigonometric model is required.

The trigonometric IRF is not monotonic. This seems counter-intuitive because one expects a monotonic relation between the ability of a person and the number of items solved. However, ability is not a deterministic cause of success on an individual item. Ability secures a good score on the average over a set of items. Thus on the level of the test as a whole, the score must grow with growing ability [23, 24]. From the point of view of the trigonometric model this is a necessary property of the data, not of the model.

Table 1. Both the Rasch and the form invariant model fulfill both measurement requirements in a sense specific for each mode

	Rasch model	**Form invariant model**
basic requirement: independence of the comparison of abilities	...from the specific set of difficulties used (specific objectivity for $R(\theta, \sigma)$)	...from the position on the scale where the measurement takes place (constancy of measure)
consequence	difference scale level (constancy of unit)	specific objectivity for $G_{\theta,\,\sigma}$

4.2 Full multidimensional Trigonometric Model

The full multidimensional trigonometric model is defined by introducing the IRF (18) into the model of Eqs. (1,2). We distinguish the parameters of the trigonometric model from those of the Rasch model by the superscript "tr". The function (18) turns the binomial model (2) into

$$q(x_{p,i}|R(\theta_p^{\rm tr}, \sigma_i^{\rm tr})) = \left[\sin^2(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4)\right]^{x_{p,i}} \left[\cos^2(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4)\right]^{1-x_{p,i}} \tag{32}$$

with the reaction function

$$R(\theta - \sigma) = \sin^2(\theta - \sigma + \pi/4) \tag{33}$$

so that the trigonometric model is

$$q^{\rm tr}(\mathbf{x}|\boldsymbol{\theta}^{\rm tr}, \boldsymbol{\sigma}^{\rm tr}) = \prod_{p=1}^{N_P}\prod_{i=1}^{N_I} \left[\sin^2(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4)\right]^{x_{p,i}} \left[\cos^2(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4)\right]^{1-x_{p,i}} . \tag{34}$$

This model does not depend solely on the scores; it depends on the individual pattern of answers given by a person; similarly it depends on the specific pattern of answers to an item. Therefore the trigonometric and logistic models quantify ability and difficulty in different ways. This is discussed in Sec. 5 below.

The trigonometric model yields the ML estimators as the solutions of the system of equations

$$\begin{aligned} 0 = \sum_{i=1}^{N_I} \Big[x_{p,i} \cot\Big(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4\Big) \\ -(1 - x_{p,i}) \tan\Big(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4\Big)\Big] , \\ p = 1 \dots, N_P ; \\ 0 = \sum_{p=1}^{N_P} \Big[x_{p,i} \cot\Big(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4\Big) \\ -(1 - x_{p,i}) \tan\Big(\theta_p^{\rm tr} - \sigma_i^{\rm tr} + \pi/4\Big)\Big] , \\ i = 1, \dots, N_I \end{aligned} \tag{35}$$

obtained by requiring the logarithmic derivatives of $q^{\rm tot}$ to vanish. In analogy to the logistic model, the convention

$$\sum_{i=1}^{N_I} \sigma_i^{\rm tr} = 0 \tag{36}$$

is used, compare Eq. (9).

The problems with ML estimation mentioned towards the end of Sec. 3 do not affect the trigonometric model. Parameters considered incidental do not jeopardize the information obtained on the remaining ones; i.e. all parameters are consistent.

In the following section we communicate quantitative results from both, the Rasch model and the trigonometric model. The data $\boldsymbol{x}$ are provided by a Monte Carlo simulation. The trigonometric estimators are discussed with respect to the information they take from the data.

5 Numerical Examples: Monte Carlo Simulations

The present section introduces a Monte Carlo simulation of a data matrix $\boldsymbol{x}$ and discusses its analysis with the Rasch model (Sec. 5.1) and the trigonometric model (Sec. 5.2). Note that the same data matrix (see next section for details of the simulation) was use for a comparative analysis by both models.

5.1 Simulated Data. Analysis with the Rasch Model

A data matrix $\boldsymbol{x}$ has been generated by a Monte-Carlo simulation of the Rasch model. The simulation comprises $N_P = 500$ persons and $N_I = 20$ items. In the simulation the true values of both parameters, θ_p and σ_i, were chosen between -4 and $+4$ at constant steps $\theta_{p+1} - \theta_p$ and constant $\sigma_{i+1} - \sigma_i$. The range $[-4, 4]$ of the parameters is encountered in practice [27]. Given the parameters, every $x_{p,i}$ was produced by a random number [25] generator according to the distribution (6).

This data matrix was subsequently analyzed with both, the Rasch model and the trigonometric model. For the Rasch analyses the code ConQuest [2] has been used. Figure 1 presents one fifth of the simulated data matrix, the part with the most competent persons. There is a black square for a correct answer $x_{p,i} = 1$ and an empty square for a false answer $x_{p,i} = 0$. The persons label the rows of the matrix; the most competent person is given by the top row, labelled by $p^{\,lg} = 1$ or $p^{\,tr} = 1$. The items label the columns; the most difficult item is given by the left hand column labelled by $i^{\,lg} = 1$ or $i^{\,tr} = 1$. The persons are ordered such that the estimators decrease with increasing $p^{\,lg} = 1$ or $p^{\,tr} = 1$. The items are ordered such that the estimators decrease with increasing $i^{\,lg}$ or $i^{\,tr}$. The estimators have been obtained by ML estimation.

On panel a) the order is defined by the estimators obtained from the Rasch model. The scores are the sufficient statistics of this model i. e. they completely determine the order in which abilities and thus persons occur; analogously for difficulties and items. The intervals of the subsequent scores $1, 2, 3, ...$ are marked by the broken lines on panel a). To be specific, there is one person $p^{\,lg} = 1$ having score 20; this

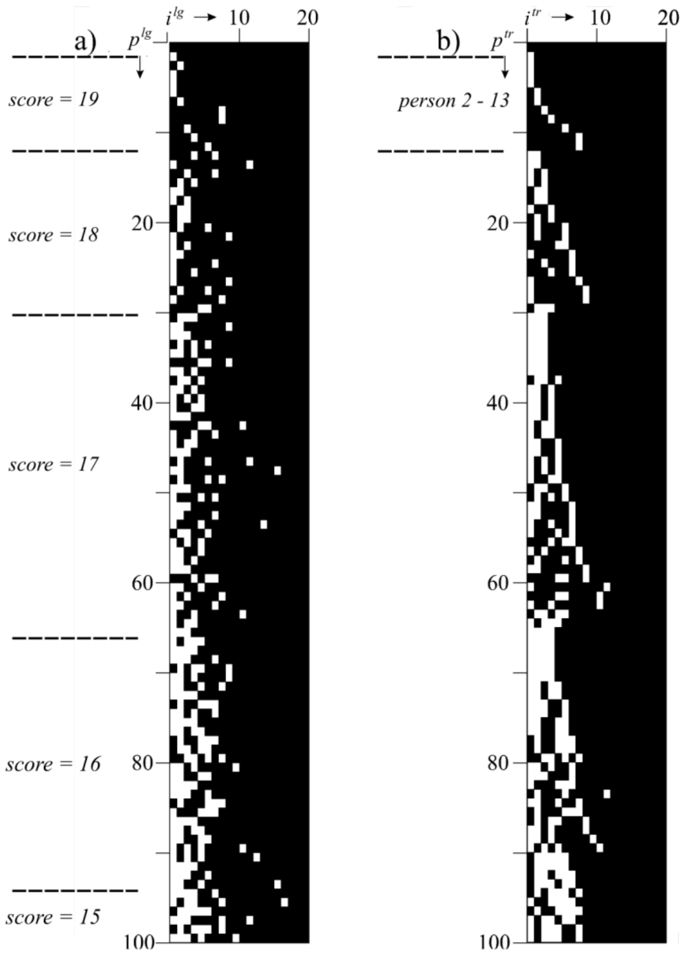

Figure 1. Data matrix given by a Monte Carlo simulation of the Rasch model (a), left) and the trigonometric model (b), right). A black (empty) square denotes a correct (false) answer. The test comprises $N_P = 500$ persons and $N_I = 20$ items. The most competent 100 persons are displayed (i.e. one fifth of the entire data matrix is shown). Both, person and item parameters, in general decrease with increasing index, score groups are separated by broken lines. On panel a) person and item parameters are given by the Rasch model; see text of Sec. 5.1. On panel b) the same data are analysed by the trigonometric model; see text of Sec. 5.2.

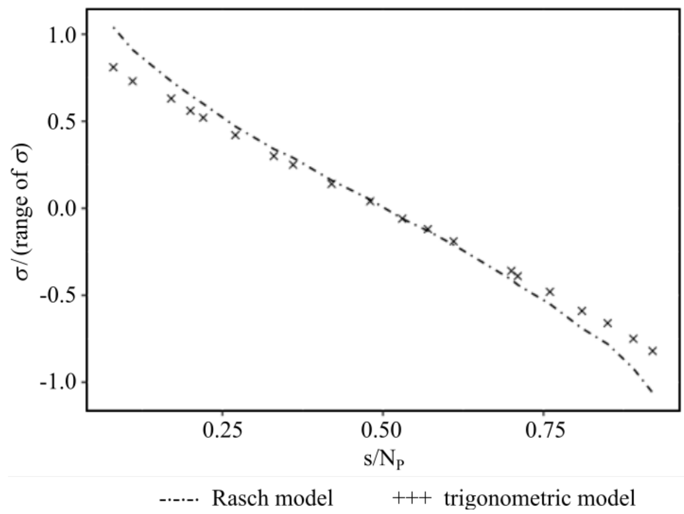

Figure 2. This figure shows item estimators derived from the data shown in Fig. 1. The estimators are plotted against the scores. Both, the estimators and the scores, have been scaled to the range of unity. The logistic estimator is given by the dashed-dotted curve and the trigonometric estimator by the crosses. See text for more details.

person has correctly answered all items. It is followed by 11 persons $p^{lg} = 2, \ldots, 12$ with score 19, followed by 17 persons $p^{lg} = 2, \ldots, 12$ etc.

We observe that there is no sharp boundary between the black area of correct answers and the white area of false answers. The probabilistic nature of the model entails a diffuse transition from one area to the other one.

Fig. 2 shows the difficulty estimator plotted against the item score for the data of Fig. 1. Both, the estimator and the score, have been scaled to the range of unity. The dashed-dotted curve gives the case of the logistic IRF. In its middle part the relation is not far from a linear function; towards the ends of the range non-linearities appear.

We turn to the trigonometric estimators derived from the same data.

5.2 Analysis with the Trigonometric Model

The data generated in Sec. 5.1 are now reanalyzed by use of the trigonometric model. The trigonometric parameters are the solutions of the system of ML equations (35). Panel b) of Fig. 1 shows the same data as panel a); however, the persons and items are now ordered according to the trigonometric estimators.

Note that the order of the persons within a group of the same score is no longer arbitrary on panel b). To recognize a systematics consider e.g. the group of persons $p^{tr} = 2, \ldots, 12$ with the score $t_p = 19$. A sequence of white holes , forming a sort of "lines", appears within the black area such that the holes move deeper and deeper into the black as p^{tr} increases. In other words: The person score remaining constant, the person parameter decreases as the unsolved item becomes easier. This visual order provided by the lines allows to interpret the additional information captured by the trigonometric parameters. A similar structure is observed for the two holes that characterize the set of persons $p^{tr} = 13, \ldots, 29$ with score $t_p = 18$ and so on. Thus the trigonometric ability estimator, although it roughly is a monotonic function of the score, is fine tuned by the difficulty of the items solved: The more difficult the solved items, the higher is the estimator. The trigonometric model unravels an order which is present in the individual patterns of the data matrix. More details are given in Sect. 4.2 of Ref. [15].

In the following, selected sections of the Monte Carlo simulation will be discussed in order to illustrate the special features of the trigonometric parameterisation. The layout and ordering of the parameters from Figure 1 are retained.

In Figure 3 the person marked with the black arrow with a score of 18 did not solve the ninth easiest item. Therefore, this person receives a lower person parameter than the persons placed above of him with a score of only 17. The same arises for the person with score 17, marked with the blue arrow in Figure 3, who also fails to solve the nineth easiest item and is

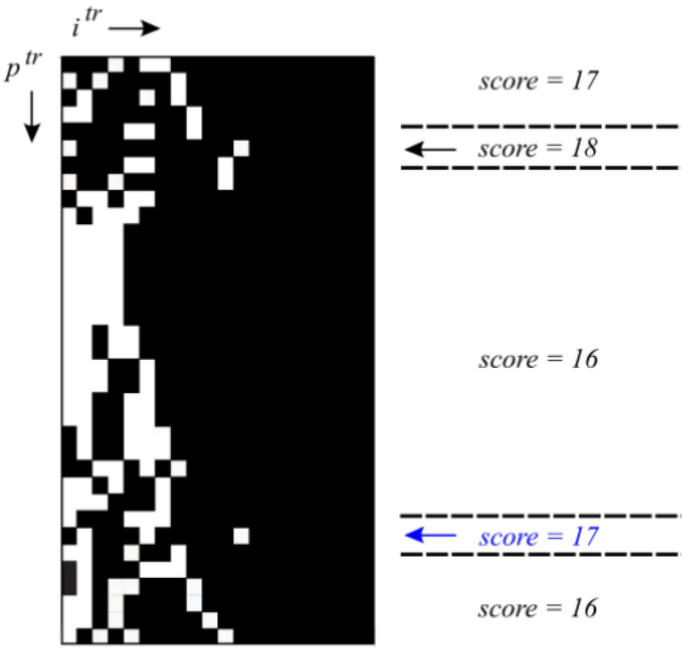

Figure 3. Enlarged section of data matrix given by the Monte Carlo simulation of the Rasch model, see Figur 1, panel b), trigonometric parametrization: here for persons with score 16, 17 and 18.

also placed below the person with score 16.

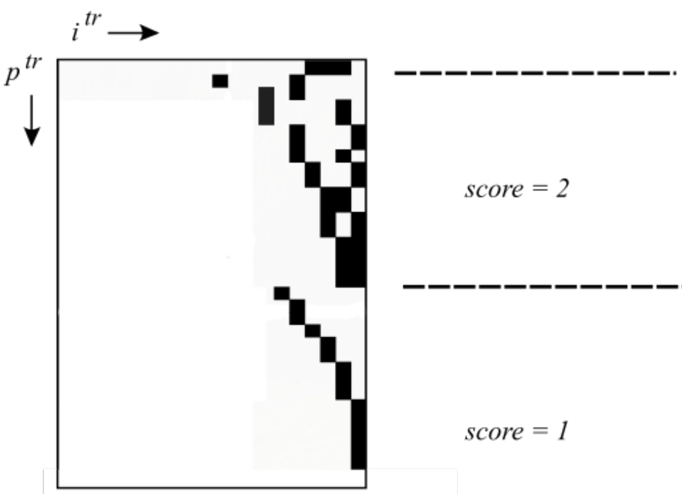

Figure 4. Enlarged section of data matrix given by the Monte Carlo simulation of the Rasch model, see Figur 1, panel b), trigonometric parametrization: here for persons with score 1 and 2.

If we now look at persons with lower scores (see Figure 4) an analogous picture emerges. The order of the persons within a group of the same score (e.g. 1) again is no longer arbitrary. A sequence of black squares, forming a sort of line (as for the white holes in Figure 1) appears. This means, that with the person score remaining constant, the person parameter decreases with the solved item becoming easier(cf. Samejima [38])

We conclude: The estimators of the trigonometric model depend on the individual patterns of answers such that it is meaningful to interpret the person parameters as ability. To which extent this is *psychologically* a more adequate model of ability remains an open question. It cannot be settled by methodological considerations alone. The item parameters could in principle be discussed likewise but we don't do this here because — in the present example as well as mostly in reality — their number N_I is much smaller than the number N_P of persons. Whence, there is more than one person for every possible person score, while the possible item scores do not all occur even once.

6 Summary and Perspectives

The present work shows that there is an alternative to derive an item-response model from a fundamental requirement about measurement: We have required form invariance, a general principle different from specific objectivity as a guide to construct statistical models. Form invariance is a symmetry, expressed by a mathematical group of transformations that connects the observations x with the parameters [16]. Form invariance is not restricted to the context of IRT. Every statistical model can be analyzed with respect to its symmetry properties and, if possible, should be made form invariant.

In order to parameterize IRT in terms of the difference between ability and difficulty, form invariance requires the measure on the parameter scale to be constant. This requirement can be expressed by group theory. In IRT, it is the IRF that sets the scale for the parameters. The group theoretical formulation of form in variance then provides the definition of the IRF.

Mathematical consequences of this IRF are the following ones: (i) The trigonometric IRF entails that the Fisher information [14] is constant. In 1980, Samejima [37] constructed IRT by requiring that the information gained on any parameter be independent of the value of its estimator. Here, we find this as a consequence of form invariance. This symmetry principle, moreover, entails constant measure in parameter space. Constant measure, in turn, is necessary if parameters are to be compared by taking their difference. (ii) The trigonometric model does not merely count the number of correct answers, it takes into account the difficulty of the items solved — although the scores roughly determine the estimators. Therefore the logistic parameters cannot be mapped onto the trigonometric ones. (iii) The trigonometric IRF is not monotonic, however on the level of the test as a whole, the average score must grow with growing ability [23, 24]. From the point of view of the trigonometric model this is a necessary property of the data, not of the model. (iv) The parameters are estimated by the principle of unconditional maximum likelihood devised by R.A. Fisher [14]. The ML equations are solvable because there are no points of instability. (v) The estimators have minimal bias. (vi) An arbitrary number of parameters can be treated as incidental without hampering inference of the remaining ones; i.e. the ML estimators are consistent.

Despite these differences, the trigonometric model also shares important properties with the logistic model. (i) It is also based on a fundamental requirement about measurement, albeit a different one than specific objectivity. (ii) It allows to conjointly estimate ability and difficulty parameters on a common scale, i.e. the estimators of ability and difficulty determine each other — either one being the instrument to measure the other one. Each one is object of measurement and instrument of measurement at the same time. (iii) Data that can be roughly ordered according to the person and item scores are successfully interpreted by the trigonometric model. I.e. the estimators basically are functions of the scores in analogy to the logistic estimators from the Rasch model. However, there is a "fine tuning" such that the trigonometric parameters pick up information from the individual patterns of answers, as in listed now.

We empirically found the following consequences of the trigonometric IRF: (i) It increases the information which the parameters pick up from the data since the estimators take care of the individual patterns of answers. The scores are not the sufficient statistics of the trigonometric model. Nevertheless we find the trigonometric estimators to be approximately given by the scores provided that the data can be roughly ordered by the scores. "Roughly ordered" means ordered except for the scatter introduced by the randomness of the data. Then the relation between score and estimator is at least as close to linear as in the logistic model. (ii) The difficulties of the solved items provide a "fine tuning" beyond the order given by the scores. This fine tuning is such that the trigonometric estimators can be interpreted as ability and difficulty: The ability estimator of a person increases if the item(s) he misses become easier, and vice versa.

For future research and application we see the following possibilities. In order to develop the theoretical understanding of the model, we want to work out formal information measures (Fisher, Shannon) and to compare them to results about the Rasch (or related) item-response models. Moreover, it seems that the degree of "noise" or randomness in the transition between incorrect and correct answers in the data matrix (i.e. the "diffuseness" between white and black areas in Figure 1) has a significance for item-response models hitherto unexplored.

While the purpose of the present contribution is to present an alternative theoretical approach to the measurement problem of IRT, the question to which extend this alternative is a psychologically adequate and useful model cannot be settled by the present methodological considerations alone. Thus we also plan to expand the analysis of simulated and real data in order to get more insight into that question. We see good reasons that the thorough foundation by a fundamental requirement about measurement presented in this work provides a good basis for theoretical and empirical progress in the sense of the above and other future perspectives.

Appendix

A Notations and abbreviations

Here, a glossary of the most important notions, conventions and notations is offered. These are given in alphabetical order and "explained" by a few key words which, however, cannot replace the definitions given in the text.

$\boldsymbol{a}$ — vector of probability amplitudes

$\boldsymbol{a}_x$ — probability amplitude for the event x to occur

$R^{\rm lg}, R^{\rm tr}$ — item response functions

form invariance — group theoretic symmetry of statistical models

IRF — item response function

IRT — item response theory

ML estimator — parameter value that maximizes the likelihood function

N_I — number of items

N_P — number of persons

q — binomial model

$q^{\rm tot}$ — dichotomous item response model

$q^{\rm lg}$ — Rasch model

$q^{\rm tr}$ — trigonometric model

s_i — item score

$\boldsymbol{\sigma}$ — vector of the item parameters

σ_i — the parameter of the i-th item

specifically objective — parameters independent of the measuring device

t_p — person score

$\boldsymbol{\theta}$ — vector of the person parameters

θ_p — the parameter of the p-th person

$\boldsymbol{x}$ — the matrix $(x_{p,i})$ of answers

$x_{p,i}$ — answer of the p-th person to the i-th item

B Proof of Equation (30)

First, we note that $G_{\pi/4} = e_1$ and $G_{\pi/4} = -e_2$, where e_1, e_2 are the basis vectors $(1,0)^T$ and $(0,1)^T$ of the vector space of the amplitude vectors $\boldsymbol{a}$. This follows by direct calculation from Eq. (19) and Eq. (20). On then infers from this and from Eq. (12) $\boldsymbol{a}(\theta + \pi/4) = G_{\theta+\pi/4} \cdot \boldsymbol{a}(0) = G_\theta G_{+\pi/4} \cdot \boldsymbol{a}(0) = G_\theta \cdot e_1$ and similarly $\boldsymbol{a}(\theta - \pi/4) = -G_\theta \cdot e_2$, which, when written in components, is Eq. (30). The derivation given here for the simplified case of G_θ with one argument according to Eq. (12) can easily be generalized to $G_{\theta,\sigma}$.

REFERENCES

[1] R. Adams, M. Wu and C. Carstensen. Application of multivariate Rasch models in international large scale educational assessment. In M. von Davier and C. H. Carstensen (Eds.) *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*, p. 271–280. New York: Springer. 2007

[2] R.J. Adams, M.L. Wu, D. Cloney and M.R. Wilson.. ACER ConQuest - Generalized Item Response Modeling Software [computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research, 2015

[3] E.B. Andersen. Conditional inference for multiple-choice questionnaires. *British J. of Mathematical and Statistical Psychology*, 26:31–44, 1973.

[4] E.B. Andersen. *Discrete Statistical Models with Social Science Applications*. North-Holland, Amsterdam, 1980.

[5] F. B. Baker, S.-H. Kim. *The Basics of Item Response Theory Using R*. Springer International Publishing, Cham, 2017. See ch. 2.3

[6] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, 53:330–418, 1763. Reprinted in Biometrika 45, 293–315 (1958) and in 'Studies in the History of Statistics and Probability' E.S. Pearson and M.G. Kendall eds., C.Griffin and Co. Ltd., London 1970 and in 'Two papers by Bayes with commentaries' W.E. Deming ed., Hafner Publishing Co., New York, 1963.

[7] J.W. Boone, J.R. Staver & M.S. Yale. *Rasch analysis in the human sciences*. Dordrecht (Netherlands): Springer. 2014

[8] M. Bühner. *Einführung in die Test- und Fragebogenkonstruktion*. Pearson Studium, München, 4nd edition, 2021. See p. 328.

[9] M. von Davier. Rasch Model. In: van der Linden, W. J. (Ed.). *Handbook of item response theory: Models, statistical tools, and applications*, Vol 1, p. 31 - 48. Boca Raton: Chapman & Hall/CRC. 2016

[10] G.H. Fischer. *Einführung in die Theorie psychologischer Tests*. Verlag Hans Huber, Bern, 1974.

[11] G.H. Fischer. Spezifische Objektivität: Eine wissenschaftstheoretische Grundlage des Rasch-Modells. In K. Kubinger, editor, *Moderne Testtheorie*, p. 87 - 111, Weinheim, 1988. Psychologie Verlagsunion.

[12] G.H. Fischer. Derivations of the Rasch model. In G.H. Fischer and I.W. Molenaar, editors, *Rasch-models: foundations, recent developments and applications*, p. 15 - 38, New York, 1995. Springer. Workshop held at the University of Vienna, Feb. 25–27, 1993.

[13] G.H. Fischer. Rasch models. In C.R Rao and S. Sinharay, editors, *Handbook of Statistics* 26 — *Psychometrics*, chapter 16, p. 515 - 585. Elsevier, Amsterdam, 2007.

[14] R.A. Fisher. Theory of statistical information. *Proc. Cambridge Philos. Soc.*, 22: p. 700 - 725, 1925. See especially p. 709.

[15] Christoph Fuhrmann. *Die trigonometrische Parametrisierung von Kompetenzen*. Springer VS, Wiesbaden, 2018.

[16] H.L. Harney. *Bayesian Inference —Parameter Estimation and Decisions*. Springer, Heidelberg, 2nd edition, 2016. Corrections and comments can be found at 'www.mpi-hd.mpg.de/harney/home'.

[17] H. Jeffreys. *Theory of Probability*, 3. ed. Oxford Classic Texts in the Physical Sciences, Oxford, 2003.

[18] H. Jeffreys. An invariant form of the prior probability in estimation problems. *Proc. of the Roy. Soc.*, 186 A:453–461, 1946.

[19] F. G. Kaiser, M. Merten & E. Wetzel. How do we know we are measuring environmental attitude? Specific objectivity as the formal validation criterion for measures of latent attributes. *Journal of Environmental Psychology*, 55, 139-146. 2018

[20] W. Kühnel. *Differential Geometry: Curves — Surfaces — Manifolds*, volume 16 of *Student Mathematical Library*. American Mathematical Society, Providence R.I., 3^{nd} edition, 2015.

[21] W.J. van der Linden (Ed.). *Handbook of item response theory: Models, statistical tools, and applications*, (vol I - III). Boca Raton: Chapman & Hall/CRC, 2016 - 2018

[22] X. Liu. *Using and developing measurement instruments in science education: A Rasch modeling approach*. Charlotte, N.C: Information Age. 2010

[23] F.M. Lord. The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13(4):517–549, 1953.

[24] F.M. Lord and M.R. Novick. *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, Mass., 1968.

[25] Microsoft, 2016. The Monte Carlo simulation has been carried out by the function RAND of EXCEL: http://support.microsoft.com/kb/828795/en-us.

[26] J. Neyman and E. L. Scott. Estimates based on partially consistent observations. *Econometrica*, 16:1–32, 1948.

[27] M. Planinic, L. Ivanjek, and A. Susak. Rasch model based analysis of the force concept inventory. *Phys. Rev. Sp. Top. – Phys. Edu. Res.*, 6:010103–1–11, 2010.

[28] M. Planinic, W.J. Boone, A. Susac & L. Ivanjek. Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 020111. 2019

[29] G. Rasch. On general laws and the meaning of measurement in psychology. In J. Neyman, editor, *Proceedings of the fourth Berkeley symposium on mathematical statistics and psychology*, pages 312–334, Berkely Ca., 1961. University of California Press.

[30] G. Rasch. An individual approach to item analysis. In P.F. Lazarsfeld and N.W. Henry, editors, *Readings in Mathematical Social Science*, pages 89–108. Massachusetts Institute of Technology, Massachusetts, 1966.

[31] G. Rasch. An informal report on a theory of objectivity in comparisons. In L.J.Th. van der Kamp and C.A.J. Vlek, editors, *Measurement theory. Proceedings of the NUFFIC international summer session in science at 'Het Oude Hof', The Hague, July 14-28, 1966*, pages 1–19, Leiden, University of Leiden, 1967.

[32] G. Rasch. A mathematical theory of objectivity and its consequences for model construction. Paper presented at the 'European Meeting on Statistics, Econometrics and Management Science', Amsterdam, The Netherlands, September 2–7, 1968.

[33] G. Rasch. Objectivitet i samfundsvidenskaberne et metodeproblem (objectivity in the social sciences as a methodological problem). *Nationaløkonomisk Tidsskrift*, 110:161–196, 1972.

[34] G. Rasch. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. In M. Blegvad (Ed.), *The Danish yearbook of Philosophy*, pages 58–94, Kopenhagen, 1977. Munksgaard.

[35] G. Rasch. *Probabilistic Models for some Intelligence and Attainment Tests*. The University of Chicago Press, Chicago, 1980.

[36] J. Rost. *Lehrbuch Testtheorie — Testkonstruktion*. Hans Huber, Bern, 2nd edition, 2004.

[37] F. Samejima. Is Bayesian estimation proper for estimating the individual's ability? Research Report 80-3, University of Tennessee, Department of Psychology, Knoxville, TN, 1980.

[38] F. Samejima. Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika*, 65(3):319–335, 2000.

[39] T.A. Warm. Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54:427–450, 1989.

[40] R. Waugh. *Applications of Rasch measurement in education*. New York: Nova Science Publishers. 2011