3

Rasch Model

Matthias von Davier

CONTENTS

3.1 Introduction

The Rasch model (Rasch, 1960, 1966a,b) is widely used and highly praised by some researchers but criticized by others who prefer more highly parameterized models for item response data. The reason behind both of these conflicting viewpoints is that the Rasch model can be derived from a small number of simple assumptions. Proponents say these assumptions form a mathematically elegant foundation of the model, while opponents say they are rather restrictive.

Even though this divide exists, many researchers in educational measurement, psychometrics, patient-reported outcomes (including quality-of-life research), and other domains are at least familiar with the name of the Rasch model and are aware that it is an important approach in the field that has inspired many applications and methodological developments.

The amount of research on the Rasch model is almost impossible to summarize in a single chapter. For more information, readers are referred to edited volumes on the Rasch model, its extensions, and diagnostics (e.g., Fischer and Molenaar, 1995; von Davier and Carstensen, 2007) or the detailed chapter by Fischer (2006). This chapter focuses on an introduction of the model, highlighting its derivation and links it to other models. In doing so, it touches only lightly on some of the most important results on specific objectivity and parameter separability (e.g., Rasch, 1960, 1966a,b), as well as on consistent estimation (e.g., Kiefer and Wolfowitz, 1956; Andersen, 1972; Haberman, 1977; de Leeuw and Verhelst, 1986; Follmann, 1988; Lindsay et al., 1991), fit diagnostics, model testing (e.g., Andersen, 1973; Molenaar, 1983; Glas, 1988a, 2007; Rost and von Davier, 1994, 1995), and attempts to put the effects of model misspecification in context (Molenaar, 1997).

3.2 Presentation of the Model

Tasks in which examinees are asked to read sentences aloud are an example of the type of items for which the Rasch model was originally developed, as is a series of problems requiring the use of essentially the same mathematical operation but with different numbers. In the motor domain, the Rasch model could be used to analyze repeated attempts of trials, such as the high jump with different heights of the bar or free throws with different distances from the basket. With the advent of the computer, we could also add repeated matches of chess players against computer programs of different strengths to this list. The commonality is that these examples confront a test taker (or athlete or player) with a series of very similar trials or items to be mastered, with success or failure expected to depend only on a specific skill of the examinee and stable common characteristics of the different items.

The following notational conventions are used: Let $\boldsymbol{U} = (U_1, \ldots, U_I)$ denote a vector of discrete observed variables, with $U_i \in \{0,1\}$ and 1 indicating success and 0 indicating failure. For each test taker $p = 1, \ldots, P$, let $U_{pi} \in \{0,1\}$ denote the test taker's response in terms of success or failure on item i. Rasch (1966a) introduces the model using the following assumptions:

1. The probability that an examinee p succeeds in trial i is

$$Pr\{U_i = 1 \mid p\} = \frac{\lambda_{pi}}{1 + \lambda_{pi}}, \tag{3.1}$$

   which is equivalent to

$$Pr\{U_i = 0 \mid p\} = \frac{1}{1 + \lambda_{pi}}. \tag{3.2}$$

2. Parameter $\lambda_{pi} = (Pr\{U_i = 1 \mid p\} \,/\, Pr\{U_i = 0 \mid p\})$ can be written as

$$\lambda_{pi} = \frac{\tau_p}{\xi_i}, \tag{3.3}$$

   which implies that $Pr\{U_i = 1 | p\} = (\tau_p/\xi_i + \tau_p)$ and $Pr\{U_i = 0 | p\} = (\xi_i/\xi_i + \tau_p)$.
3. Stochastic independence of these probabilities holds for examinee p attempting multiple trials $i = 1, \ldots, I$.

Note that with the transformation $\theta_p = \ln(\tau_p)$ and $\beta_i = \ln(\xi_i)$, we obtain an equivalent form of the Rasch model:

$$Pr\{U_i = u \mid p\} = \frac{\exp(u(\theta_p - \beta_i))}{1 + \exp(\theta_p - \beta_i)}, \tag{3.4}$$

which relates more readily to other logistic models (Volume One, Chapter 2).

Parameters τ_p and ξ_i are unknown quantities that describe the level of skill or proficiency of examinee p and the difficulty of item i, respectively. Presenting the Rasch model as in Equations 3.1, 3.2, or 3.3 may be less familiar but helps to relate the model to a wider range of approaches for the analysis of (multivariate) binary data. Also, the fractions in Equation 3.3 relate the Rasch model in a straightforward way to the simple estimate of success probabilities based on sample frequency of successes and failures for repeatable trials. Further, note that the ratio of parameters τ_p and ξ_i in the second assumption are the (betting) odds for a binary variable, that is, the ratio of the probabilities of success and failure. A few very important properties of the Rasch model (as well as some of more general item response theory, or IRT, models) can conveniently be illustrated based on this representation of the model. If the Rasch model holds for all examinees p in a given population and for items $i = 1, \ldots, I$, it can be shown that

- Comparisons of examinees can be carried out independent of the items involved. For any item i, the ratio of parameters for two examinees p and p' equals

$$\frac{\lambda_{ip}}{\lambda_{ip'}} = \frac{\tau_p}{\tau_{p'}}$$

  independently of which item i was chosen.
- Similarly, comparisons of items i and i' can be carried out independently of the examinees involved. For any examinee p, we have

$$\frac{\lambda_{ip}}{\lambda_{i'p}} = \frac{\xi_{i'}}{\xi_i}.$$

- Likewise, a direct comparison of item difficulties or examinee proficiencies is possible even when two sets of observations overlap only partially, or if two sets of observations are connected only through a third set of observations (items or examinees). For example, if examinee p took only items i and i' while examinee q took items i' and i'', how do we compare between item i and i''? The answer is easily obtained if the different λ_{ip} parameters are known. This follows because

$$\frac{(\lambda_{ip}/\lambda_{i'p})}{(\lambda_{i''q}/\lambda_{i'q})} = \frac{(\xi_{i'}/\xi_i)}{(\xi_{i'}/\xi_{i''})} = \frac{\xi_{i''}}{\xi_i},$$

  even though the two items have never been taken together, provided that there is a item i' taken by two (groups of) examinees that links items i and i''. This feature can be utilized for comparisons among items or test takers in the same way. As an example, Glas (1988b) shows how the Rasch model can be applied to multistage testing with a large number of different test forms assigned to different examinees.

There is only one issue: The parameters λ_{ij} are unknown, and naïve estimators are not available because each test-taker-by-item interaction is, in a typical testing situation,

observed only once. The solution to this problem is an approach that capitalizes on the sufficiency (parameter separability) properties of the Rasch model. For any two items i and i', consider the conditional probability of solving i given that exactly one of the two items was solved:

$$\Pr\{U_{ip} = 1 \mid U_{ip} + U_{i'p} = 1\}.$$

If the Rasch model holds, local stochastic independence can be used. Then, applying Equation 3.3, the probability can be written as

$$\frac{\Pr\{U_{ip} = 1\}\Pr\{U_{i'p} = 0\}}{\Pr\{U_{ip} = 1\}\Pr\{U_{i'p} = 0\} + \Pr\{U_{ip} = 0\}\Pr\{U_{i'p} = 1\}} = \frac{(\tau_p/\xi_i + \tau_p)(\xi_{i'}/\xi_{i'} + \tau_p)}{(\tau_p/\xi_i + \tau_p)(\xi_{i'}/\xi_{i'} + \tau_p) + (\xi_{i*}/\xi_i + \tau_p)(\tau_p/\xi_{i'} + \tau_p)}$$

which, upon some simplification, yields

$$\Pr\{U_{ip} = 1 \mid U_{ip} + U_{i'p} = 1\} = \frac{\xi_{i'}}{\xi_{i'} + \xi_i} \tag{3.5}$$

for any examinee p. This is a central result, which we will discuss for the case of more than two items below. For now, we only note that Equation 3.5 shows that by conditioning on the total score on any two items, the examinee parameter τ_p is effectively eliminated from the conditional probability.

Now assume that $p = 1, \ldots, P$ test takers with parameters τ_p were sampled randomly from the population for which the Rasch model holds. Consider the sample estimate of the marginal conditional probability of observing success on item i given that exactly one of the two items i and i' was solved:

$$\widehat{\Pr}\{U_i = 1 \mid U_i + U_{i'} = 1\} = \frac{1}{P}\sum_{p=1}^{P} 1_{\{U_{ip} = 1 \mid U_{ip} + U_{i'p} = 1\}}. \tag{3.6}$$

Taking the expectation yields

$$E\left(\frac{1}{P}\sum_{p=1}^{P} 1_{\{U_{ip} = 1 \mid U_{ip} + U_{i'p} = 1\}}\right) = \sum_{p=1}^{P} \frac{1}{P} E(1_{\{U_{ip} = 1 \mid U_{ip} + U_{i'p} = 1\}}).$$

Inserting Equation 3.4 for each of the expected values inside the sum and simplifying yields

$$E(\widehat{\Pr}\{U_i = 1 \mid U_i + U_{i'} = 1\}) = \frac{\xi_{i'}}{\xi_{i'} + \xi_i}. \tag{3.7}$$

Note that this is the expected value of the sample estimate of the marginal conditional probability $\widehat{\Pr}\{U_i = 1 \mid U_i + U_{i'} = 1\}$ of observing success on item i given that exactly one of

the two items i and i' was solved. By symmetry, the ratio of this estimator and its complement yields

$$\frac{\widehat{\Pr}\{U_i = 1 \mid U_i + U_{i'} = 1\}}{\widehat{\Pr}\{U_{i'} = 1 \mid U_i + U_{i'} = 1\}} \approx \frac{\xi_{i'}}{\xi_i}. \tag{3.8}$$

This means that the ratio of any two item parameters can be compared by marginal two-way counts independent of the sample of test takers that was used.

As Rasch (1966a) stated:

> Our aim is to develop probabilistic models in the application of which the population can be ignored. It was a discovery of some mathematical significance that such models could be constructed, and it seemed remarkable that data collected in routine psychological testing could be fairly well represented by such models (p. 89).

The results presented in Equations 3.7 and 3.8 allow comparisons of item difficulties without looking at properties of the sample of examinees that were used to obtain the sample statistics in Equation 3.6. In a testing situation, we typically have many more test takers than items in the test. These results are particularly useful in this situation: If the Rasch model holds, the only requirement to be fulfilled is a large enough sample size from the population of interest in order to obtain estimates of desired accuracy for Equations 3.6 and 3.8. By symmetry, in a situation in which there are few test takers but a large number of randomly chosen items, examinees can be compared independent of the sample of items they took.

This particular feature of the Rasch model is sometimes referred to as "parameter separability" or "specific objectivity." The same feature is not found in most other higher parametric item response models, with only a few exceptions (e.g., Verhelst and Glas, 1995). The feature, which is more generally based on the fact that the Rasch model is an exponential family model with simple sufficient statistics, is important in derivations of conditional maximum likelihood (CML) estimation methods that can be established simultaneously for all item parameters without introducing bias by nuisance parameters or assumptions about the distribution of unobserved test-taker parameters (e.g., Andersen, 1972; Verhelst and Glas, 1995). The feature can also be used to estimate pairwise relationships between item parameters (Zwinderman, 1995). Moreover, this feature puts the Rasch model—or better, its ability to compare two items (or two test takers) independent of the others—in the context of exponential family models for paired comparisons in incomplete designs (Zermelo, 1929; Bradley and Terry, 1952). The connection to paired comparison models also allows stating conditions for the existence of maximum likelihood estimators (Fischer, 1981). It is of particular importance for linking or equating multiple test forms using the Rasch model as it allows comparisons across a large set of items or test takers without the requirement that all test takers have responded to all the items, for example with multistage tests (Glas, 1988b).

3.3 Parameter Estimation

Several estimation methods are available for the Rasch model. This chapter focuses on maximum likelihood methods and presents a short description of each method while

providing more detail on one that can be considered as most directly utilizing the mathematical properties of the Rasch model. Commonly discussed estimation methods for the Rasch model are

1. Joint maximum likelihood (JML): This method attempts to estimate all parameters τ_p and ξ_i for items $i = 1, \ldots, I$ and respondents $p = 1, \ldots, P$ simultaneously. There are well-known results that show that a joint estimation of parameters leads to biased estimates if $I \ll P$ (Kiefer and Wolfowitz, 1956). For the Rasch model, the main result given by Andersen (1972) and Haberman (1977) is that the bias is of order $(I/(I - 1))$. Haberman (1977, 2004) notes that besides this bias of JML estimates, JML estimation is often problematic because (finite) estimates do not exist (see also Fischer, 1981) when there are respondents with perfect scores or with a vector of all incorrect responses. It should be noted that alternatives exist that do not share these issues, so there is no need to use JML methods in typical testing situations.

2. Maximum marginal likelihood (MML): This method is, among maximum likelihood methods, the most commonly used approach for higher parameterized item response models. In order to apply MML estimation, an assumption concerning the distribution $\phi(\theta)$ of examinee parameters is necessary to enable the calculation of the marginal probability:

$$Pr\{U_i = 1\} = \int_{-\infty}^{+\infty} Pr\{U_i = 1 \mid \theta\}\phi(\theta)d\theta. \tag{3.9}$$

   While MML is a commonly used approach in IRT also available for the Rasch model (Thissen, 1982), there is an issue of choice of the distribution $\phi(\theta)$. For example, many software packages allow only normal distributions here and also approximate the integral in the equation using numerical quadrature. Note, however, that the distribution does not need to be normal (Andersen and Madsen, 1977; Thissen, 1982) and that the MML approach can be used with more general distributions (e.g., Xu and von Davier, 2008; Xu and Jia, 2011).

3. Pairwise (conditional) estimation: As shown in Equations 3.7 and 3.8, simple ratios of probabilities can be used to eliminate the parameters associated with respondents to obtain estimates of parameters τ_p and ξ_i (e.g., Zwinderman, 1995; Andrich, 2001).

4. Nonparametric or semi-nonparametric estimation using located latent classes (LLCA): de Leeuw and Verhelst (1986), Lindsay, Clogg and Grego (1991), Follmann (1988), and Formann (1994) describe an estimation method that is based on approximating the integral in Equation 3.9 by a sum over a small number of unobserved groups instead. This approach estimates the marginal probability as a finite sum:

$$Pr\{U_i = 1\} = \sum_{g=1}^{G} Pr\{U_i = 1 \mid \theta_g\}\pi_g \tag{3.10}$$

   over $g = 1 \ldots G$ latent groups with locations θ_g and relative sizes π_g with $\sum_g \pi_g = 1$. Heinen (1996) describes these approaches for the Rasch model in more detail, and

Haberman et al. (2008) describe an extension of this approach to multidimensional item response models, while Kelderman (1984) describes log-linear-type Rasch models.

5. CML: This is the estimation approach that can be considered most fundamentally based on the existence of simple sufficient statistics (i.e., what Rasch called specific objectivity) in the Rasch model. That is, the CML approach is based on a general version of result presented in Equation 3.5 capitalizing on the fact that the sum of correct responses on the number of items $i = 1, ..., I$ to which an examinee p responds is a sufficient statistic for his or her parameter θ_p (e.g., Andersen, 1977; Bickel and Doksum, 1977). Consequently, utilizing the factorization theorem of sufficiency allows the construction of estimation equations for $\beta_1, ..., \beta_I$ that do not contain any examinee parameters. Fischer (1981) provides necessary and sufficient conditions for the existence and uniqueness of CML estimates in the Rasch model, while Eggen (2000) describes minimal loss of information due to the conditioning on this sufficient statistic in CML relative to JML and MML. It is important to note that, unlike the approach in Equations 3.9 and 3.10, CML allows item parameter estimation without making any assumption about the distribution of θ.

The remainder of this section presents the CML estimation method in more detail. For the Rasch model, using local stochastic independence, the probability of a vector of responses $\boldsymbol{u} = (u_1, ..., u_I)$ is

$$\Pr\{\boldsymbol{u} \mid p\} = \prod_{i=1}^{I} \frac{\exp(u_i(\theta_p - \beta_i))}{1 + \exp(\theta_p - \beta_i)}. \tag{3.11}$$

Let $S = \sum_{i=1}^{I} u_i$ denote the sum of correct responses for the vector. Then, we have

$$\Pr\{\boldsymbol{u} \mid p\} = \exp\{S\theta_p\} \prod_{i=1}^{I} \frac{\exp(-\beta_i u_i)}{1 + \exp(\theta_p - \beta_i)}. \tag{3.12}$$

Further, let $\Pr\{S \mid p\}$ denote the probability of score S when testing examinee p, which can be written as

$$\Pr\{S \mid p\} = \sum_{\left\{(v_1, ..., v_I): \sum_{i=1}^{I} v_i = S\right\}} \exp(S\theta_p) \prod_{i=1}^{I} \frac{\exp(-\beta_i v_i)}{1 + \exp(\theta_p - \beta_i)}. \tag{3.13}$$

The conditional probability of the response vector given the sum score S is defined as

$$\Pr\{\boldsymbol{u} \mid S, p\} = \frac{\Pr\{u_1, ..., u_I \mid p\}}{\Pr\{S \mid p\}}. \tag{3.14}$$

Inserting Equations 3.12 and 3.13 into Equation 3.14 and rearranging the result, we obtain

$$\Pr\{\boldsymbol{u} \,|\, S, p\} = \frac{\exp\left(\sum_{i=1}^{I} -\beta_i u_i\right)}{\sum_{\left\{v:\sum_{i=1}^{I} v_i = S\right\}} \exp\left(\sum_{i=1}^{I} -\beta_i v_i\right)}, \tag{3.15}$$

which does not contain the examinee parameter θ_p. Introducing the symmetric functions (e.g., Andersen, 1972; Gustafsson, 1980)

$$\gamma_S(\boldsymbol{\beta}) = \sum_{\left\{v:\sum_{i=1}^{I} v_i = S\right\}} \exp\left(-\sum_{i=1}^{I} \beta_i v_i\right) \tag{3.16}$$

and marginalizing over the sample distribution of the examinees, we obtain

$$\Pr\{\boldsymbol{u} \,|\, S\} = \frac{\exp\left(-\sum_{i=1}^{I} \beta_i u_i\right)}{\gamma_S(\boldsymbol{\beta})} \tag{3.17}$$

for the conditional probability of a response vector $\boldsymbol{u} = (u_1, \ldots, u_I)$ given total score S. Note that $\Pr\{0, \ldots, 0|0\} = \Pr\{1, \ldots, 1|I\} = 1$ holds for the extreme score groups 0 and I because there is a unique single pattern of responses that produces each of these scores.

It is important to note that all items are not required to be taken by all respondents: The same approach to conditional estimation can be applied using a generalized version of the symmetric functions and Equation 3.17 in which an indicator vector $J_p = (j_{p1}, \ldots, j_{pI})$, with $j_{pi} \in \{0, 1\}$ associated with each test taker p showing the subset of items taken by this person. We can then define

$$\Pr\{\boldsymbol{u} \,|\, S_{J_p}, p\} = \frac{\exp\left(\sum_{i=1}^{I} -\beta_i j_{pi} u_i\right)}{\sum_{\left\{v:\sum_{i=1}^{I} v_i = S_{J_p}\right\}} \exp\left(\sum_{i=1}^{I} -\beta_i j_{pi} v_i\right)} \tag{3.18}$$

with $S_{J_p} = \sum_{i=1}^{I} j_{pi} u_i$ defining the total score based on the subset of items actually taken by p. All subsequent calculations can be adjusted similarly to realize this more general approach, with the advantage of, for instance, estimation of all parameters across multiple test forms in incomplete block designs on the same scale. Minimal requirements for this type of model-based linking were published as early as 1929 by Zermelo.

Note that the symmetric functions in Equation 3.16 can be written in a recursive manner as

$$\gamma_S(\boldsymbol{\beta}) = \exp(-\beta_i)\gamma^{(i)}_{S-1}(\boldsymbol{\beta}) + \gamma^{(i)}_S(\boldsymbol{\beta}), \tag{3.19}$$

where $\gamma_s^{(i)}(\boldsymbol{\beta})$ is defined as the symmetric function calculated under omission of item i (e.g., Andersen, 1973; Gustafsson, 1980; van der Linden, Volume Two, Chapter 6). Moreover, application of some basic calculus shows that the derivative of $\gamma_S(\boldsymbol{\beta})$ with respect to β_i is

$$\frac{\partial \gamma_S(\boldsymbol{\beta})}{\partial \beta_i} = -\exp(-\beta_i)\gamma^{(i)}{}_{S-1}(\boldsymbol{\beta}). \tag{3.20}$$

The conditional likelihood function for $p = 1, \ldots, P$ independently observed respondents, each responding to $i = 1, \ldots, I$ items, can be written as

$$\ln L_M(\boldsymbol{\beta} \mid \boldsymbol{u}_1, \ldots, \boldsymbol{u}_P) = \ln\left(\prod_{p=1}^{P}\left[\sum_T \Pr\{\boldsymbol{u_p} \mid T\}\pi_T \right] \right) = \sum_{p=1}^{P} \ln(\Pr\{\boldsymbol{u_p} \mid S_p\}\pi_{S_p}) \tag{3.21}$$

with score probabilities $\Pr\{S\} = \pi_S$ that can be based on relative sample frequencies H_s/P of the examinees' total scores $S_p = 0, \ldots, I$. The right-hand side of Equation 3.21 holds because $\Pr\{\boldsymbol{u}|T\} = 0$ if $T \neq S_p$. Dropping the terms $\ln\pi_S$ that do not contain any item parameters and inserting Equation 3.17 into Equation 3.21 yields

$$\ln L(\boldsymbol{\beta}|\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P) = \sum_{p=1}^{P} \ln\left[\frac{\exp\left(-\sum_{i=1}^{I} \beta_i u_{pi} \right)}{\gamma_{S_p}(\boldsymbol{\beta})} \right]. \tag{3.22}$$

Collecting the terms in Equation 3.22 into score groups $S = 1, \ldots, I-1$ (the terms for the extreme score groups vanish), we obtain

$$\ln L(\boldsymbol{\beta}|\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P) = \sum_{S=1}^{I-1} H_S[-\ln \gamma_S(\boldsymbol{\beta})] - \left[\sum_{i=1}^{I} \beta_i F_i \right] \tag{3.23}$$

with F_i denoting the count of the correct responses for item i in score groups $S = 1, \ldots, I-1$, and H_S denoting the frequency of score S in the sample.

The partial derivatives of Equation 3.23 with respect to β_i are

$$\frac{\partial \ln L(\boldsymbol{\beta}|\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P)}{\partial \beta_i} = \left[-\sum_{S=1}^{I-1} H_S \frac{\partial \ln \gamma_S(\boldsymbol{\beta})}{\partial \beta_i} \right] - F_i, \tag{3.24}$$

which upon transformation, simplification, and insertion of Equation 3.19 into Equation 3.24 yield

$$\frac{\partial \ln L(\boldsymbol{\beta}|\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P)}{\partial \beta_i} = \left[\sum_{S=1}^{I-1} H_S \frac{\exp(-\beta_i)\gamma^{(i)}{}_{S-1}(\boldsymbol{\beta})}{\gamma_S(\boldsymbol{\beta})} \right] - F_i. \tag{3.25}$$

Note that this is a result to be expected for exponential families (Bickel and Doksum, 1977; Andersen, 1980; Haberman, Volume Two, Chapter 4) because

$$\Pr\{U_i = 1 \mid S\} = \frac{\exp(-\beta_i)\gamma^{(i)}_{S-1}(\boldsymbol{\beta})}{\gamma_S(\boldsymbol{\beta})} \tag{3.26}$$

holds for the conditional form of the Rasch model. The derivatives in Equation 3.25 can be used either for the Newton–Raphson algorithm or with gradient (von Davier and Rost, 1995) or quasi-Newton methods (e.g., Dennis and Schnabel, 1996) to obtain maximum likelihood estimates of $\beta_1, \ldots, \beta_I$. When Newton methods are used, the second-order derivatives of Equation 3.25 are easily derived. Solving Equation 3.19 for $\exp(-\beta_i)\gamma^{(i)}_{S-1}(\boldsymbol{\beta})$ and inserting the result into Equation 3.26 yields

$$\frac{\gamma_S(\boldsymbol{\beta}) - \gamma^{(i)}_S(\boldsymbol{\beta})}{\gamma_S(\boldsymbol{\beta})} = \frac{\exp(-\beta_i)\gamma^{(i)}_{S-1}(\boldsymbol{\beta})}{\gamma_S(\boldsymbol{\beta})} = \Pr\{U_i = 1 \mid S\},$$

which implies

$$\frac{\gamma^{(i)}_S(\boldsymbol{\beta})}{\gamma_S(\boldsymbol{\beta})} = 1 - \Pr\{U_i = 1 \mid S\}.$$

Using the quotient rule for differentiation, along with some collecting of the terms and further simplifying, provides

$$\frac{\partial L(\boldsymbol{\beta} \mid u_1, \ldots, u_P)}{\partial^2 \beta_i} = \sum_{S=1}^{I-1} H_S \left[\frac{\left(\gamma_S(\boldsymbol{\beta}) - \gamma^{(i)}_S(\boldsymbol{\beta})\right)\left(\gamma^{(i)}_S(\boldsymbol{\beta})\right)}{[\gamma_S(\boldsymbol{\beta})]^2} \right]. \tag{3.27}$$

Alternatively, applying some of the properties of exponential families to the Rasch model (Bickel and Doksum, 1977; Haberman, 1977; Andersen, 1980) in order to find the second-order derivative of Equation 3.24 with respect to β_i will produce the equivalent result

$$\frac{\partial L(\boldsymbol{\beta} \mid u_1, \ldots, u_P)}{\partial^2 \beta_i} = \left[\sum_{S=1}^{I-1} H_S \Pr\{U_i = 1 \mid S\}(1 - \Pr\{U_i = 1 \mid S\}) \right]. \tag{3.28}$$

Once item parameter estimates $\hat{\beta}_1, \ldots, \hat{\beta}_I$ are available, they can be plugged into the joint likelihood equations. Solving the equations while holding $\hat{\beta}_1, \ldots, \hat{\beta}_I$ constant provides estimates of the examinee parameters. Alternatively, bias-corrected estimators can be found by maximizing a penalized likelihood function while holding the item parameters fixed (Warm, 1989; Firth, 1993). Finally, Bayesian estimates of the examinee parameters can be generated using their posterior modes or expectations, for instance, by adopting some "weakly informative" (Gelman et al., 2003; Gelman, 2007) prior distributions for them.

3.4 Model Fit

Numerous approaches to test the fit of the Rasch model are available (e.g., Andersen, 1973; Martin-Löf, 1973; van den Wollenberg, 1982; Molenaar, 1983; Glas, 1988a). Only a selection of tests specifically developed for this particular model will be discussed. More recent fit measures based on generalizations of Pearson χ^2 statistics have been developed for the Rasch model by Glas (1988a, 2007) and Glas and Verhelst (1995). For a more comprehensive overview of these methods, refer to Glas (Volume Two, Chapter 10).

Many of the early attempts to test the fit of the Rasch model were based on the fact that if the model holds, its item parameters can be estimated consistently regardless of the population involved (Rasch, 1966a; Andersen, 1973). These tests typically split the sample into two or more groups and estimate the parameters of the conditional model in each of them separately. The rationale behind these tests is that if parameter estimates differ substantially in different subsamples, then the Rasch model cannot adequately describe the whole sample and hence has to be rejected for the population from which the sample was drawn. Rasch (1960) utilized a graphical representation in the form of a simple scatterplot to determine whether separately estimated parameters line up without any obvious deviations. This approach is sometimes referred to as Rasch's graphical model test.

Andersen (1973) proposed a more rigorous statistical test of the Rasch model. This approach requires splitting the sample into groups of respondents according to their sum score S and estimating the Rasch model in each of the (nonextreme score) groups separately. Note that CML estimation is possible even if all respondents have the same (nonextreme) sum score, because the likelihood function in Equation 3.23 maximized with respect to the item parameters does not depend on the actual total scores observed. If we split the sample by score groups, we may define

$$\ln L_S(\boldsymbol{\beta}|\boldsymbol{u}'_1,\ldots,\boldsymbol{u}'_{H_s}) = H_S[-\ln \gamma_S(\boldsymbol{\beta})] - \left[\sum_{i=1}^{I} \beta_i F_{iS} \right], \tag{3.29}$$

where $F_{iS} = \sum_{p|S} u_{pi}$ denotes the count of the correct responses for item i only for respondents with sum score S and H_S is defined as above.

Maximizing $\ln L_S(\boldsymbol{\beta}|\boldsymbol{u}'_1,\ldots,\boldsymbol{u}'_{H_s})$ with respect to the item parameters for each score group $S = 1, \ldots, I-1$ separately yields score-group-based estimates $\hat{\boldsymbol{\beta}}_s = (\hat{\beta}_{1s},\ldots,\hat{\beta}_{Is})$. As a test for the Rasch model, Andersen (1973) proposed evaluation of the likelihood ratio

$$LR_{Andersen} = \frac{L_S(\hat{\boldsymbol{\beta}}|\boldsymbol{u}_1,\ldots,\boldsymbol{u}_P)}{\prod_{S=1}^{I-1} L_S(\hat{\boldsymbol{\beta}}_s|\boldsymbol{u}'_1,\ldots,\boldsymbol{u}'_{H_s})}. \tag{3.30}$$

This expression evaluates the gain in likelihood based on separate estimation of item parameters in score groups, $\hat{\boldsymbol{\beta}}_s$, against single estimates based on the total sample, $\hat{\boldsymbol{\beta}}$.

Alternatively, the model can be tested splitting the sample into groups using background variables, such as level of education (e.g., years of schooling in broad categories) or gender. Rost and von Davier (Rost and von Davier, 1995; Volume One, Chapter 23) proposed using a mixture distribution Rasch model as the alternative hypothesis in a test of parameter invariance, introducing a split of the sample of examinees into latent groups.

Another test of the Rasch model is the Martin-Löf test (e.g., Glas and Verhelst, 1995; Verguts and De Boeck, 2000). This test is one of the few that checks on the unidimensionality assumption for the Rasch model (Verhelst, 2001). Unlike the previous tests, which split the examinee sample into subsamples to check on its homogeneity, the Martin-Löf test splits the item set into two subsets and then addresses their dimensionality.

Some other fit measures discussed in the literature are based on response residuals:

$$r_{pi} = u_{pi} - \Pr\{U_i = 1 \mid p\}.$$

These residuals are aggregated either by first squaring and then adding them or by first adding them, and then squaring the result. Aggregation across items yields a measure of person fit; across persons, it yields a measure of item fit. Depending on the type of aggregation, different types of standardization have been proposed (Wright and Stone, 1979); however, the exact null distribution of these standardized statistics is unknown (Wells and Hambleton, Volume Two, Chapter 20).

Fit measures based on CML inference have also been presented (e.g., Rost and von Davier, 1994). More recent approaches to model fit include the work by von Davier and Molenaar (2003) on person fit indices for Rasch models and its extensions to multiple populations and mixture distributions. Klauer (1991) developed a uniformly most powerful test for the person parameter in the Rasch model, while Ponocny (2000) presented an exact person fit index. Verhelst (2001) compared the Martin-Löf test with van den Wollenberg's (1982) as well as Molenaar's (1983) splitter item technique. He also proposed a new test statistic that addresses some of the statistical issues involved with the latter.

An important issue is the effect of misfit of the Rasch model on its use. Molenaar (1997) outlined a rationale to evaluate whether the increase in model complexity involved in replacing the Rasch model by a more highly parametrized model is justified. Also, person parameters estimates will still correlate highly, often substantially above 0.9, between different unidimensional IRT models applied to the same data. Finally, if the Rasch model is found not to fit the data well, more general models are available that preserve many of its characteristics, including the one-parameter logistic model (OPLM) (Verhelst and Glas, 1995) and logistic mixture models (Rost and von Davier, 1995; Volume One, Chapter 23).

3.5 Empirical Example

This empirical example is based on 10 items taken from a cognitive skills test for children ages 9–13. The dataset is available through the exercises in the textbook by Rost (2004). In the current example, the focus is on parameter recovery using different estimation methods. The item parameter estimates from the original dataset were treated as the generating parameters for a simulated dataset with known characteristics. These generating parameters (in Table 3.1 denoted by "Truth") served as the targets that were to be recovered by the different estimation methods. The dataset was simulated with 10 item responses for each of 15,000 test takers drawn from a normal ability distribution with mean 0.0 and standard deviation 1.5.

The item parameters were estimated utilizing the CML and JML estimation methods in WINMIRA 2001 (von Davier, 1994, 2001) and BIGSTEPS (Wright and Linacre, 1991),

TABLE 3.1

Generating Parameters (Truth) and CML, LLCA, and JML Item Parameter Estimates for the Example Data

Item	Truth	CML	S.E.	LLCA	S.E.	JML	S.E.
1	−0.941	−0.921	0.021	−0.921	0.020	−1.030	0.020
2	−0.457	−0.461	0.020	−0.462	0.020	−0.520	0.020
3	0.265	0.241	0.020	0.242	0.019	0.270	0.020
4	0.919	0.917	0.021	0.917	0.020	1.030	0.020
5	1.335	1.350	0.022	1.351	0.021	1.520	0.020
6	−1.223	−1.235	0.022	−1.235	0.021	−1.390	0.020
7	0.022	0.033	0.020	0.033	0.019	0.040	0.020
8	−0.457	−0.429	0.020	−0.430	0.020	−0.480	0.020
9	0.832	0.816	0.021	0.817	0.020	0.910	0.020
10	−0.295	−0.310	0.020	−0.311	0.020	−0.350	0.020

respectively, and a semiparametric approach (Heinen, 1996). The latter was implemented specifying four different latent classes (LLCA) and estimating the model with the mdltm software (von Davier, 2008).

The comparison focused primarily on the differences among the CML, JML, and LLCA methods for the estimation of item parameters and between JML and weighted likelihood estimation (WLE; Warm, 1989) for the estimation of the person parameters. Table 3.1 shows the generating item parameters and their CML, LLCA, and JML estimates along with the standard errors as provided by the programs.

The estimated standard errors for the three methods were comparable, while the estimates for CML and LLCA were virtually identical and appeared to be closer to the true values of the parameters than the JML parameter estimates. Since CML and LLCA are virtually identical, we focus our discussion and compare the CML and JML estimates only.

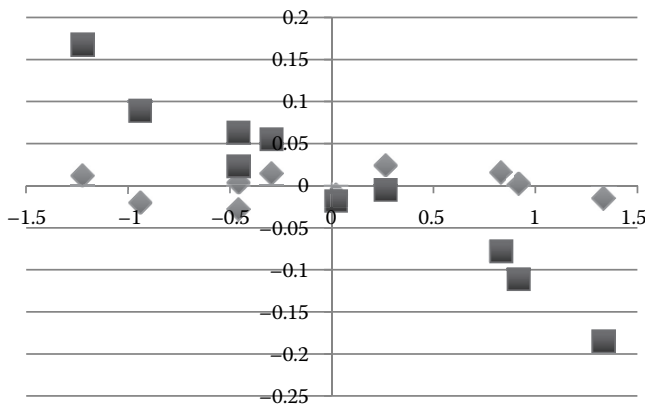Figure 3.1 shows the differences between the CML and JML estimates and the true item parameters.

FIGURE 3.1

Differences between CML and true parameters (diamonds) and JML and true parameters (squares) for the dataset.

TABLE 3.2

Person Parameter Estimates Obtained Using Three Different Methods

Score	Frequency	MLE	S.E.	WLE	S.E.	JMLE	S.E.
0	780			−3.348	1.564	(−3.27)	(1.47)
1	1162	−2.429	1.079	−2.094	0.961	−2.49	1.08
2	1493	−1.565	0.824	−1.415	0.795	−1.61	0.83
3	1556	−0.970	0.730	−0.894	0.721	−1.00	0.74
4	1637	−0.472	0.689	−0.440	0.687	−0.49	0.70
5	1702	−0.007	0.678	−0.009	0.678	−0.01	0.69
6	1621	0.459	0.692	0.424	0.690	0.47	0.70
7	1591	0.964	0.735	0.886	0.726	0.99	0.75
8	1510	1.567	0.830	1.417	0.801	1.61	0.84
9	1198	2.442	1.084	2.109	0.969	2.50	1.09
10	750			3.377	1.574	(3.29)	(1.48)

Note: JML estimates and standard errors for the extreme-score groups were not estimated but extrapolated.

As Figure 3.1 shows, the bias of JML estimates correlated negatively with the true parameter, while much smaller differences were seen for CML with no visible correlation with the true parameter values.

Table 3.2 shows the person parameter estimates obtained using subsequent maximum likelihood estimation (MLE) with the CML item difficulties as fixed constants (MLE), WLE, and the estimates when jointly maximizing the item and person parameters (joint maximum likelihood estimation, or JMLE).

While it is possible to calculate means and variances of the JML and ML estimates across score groups 1–9, it is questionable whether anything is gained given that 1530 cases in the data were found to be in extreme score groups for which (finite) estimates do not exist (Haberman, 1977, 2004; Fischer, 1981). Also, note that the entries in the table for JML and score groups 0 and 10 were extrapolated—not estimated—values, so it is meaningless to evaluate their average and variance as well. The mean estimate for the person distribution for the LLCA approach was 0.001, with a standard deviation of 1.423. The average of the WLEs was 0.001 and their standard deviation was 1.567, respectively. These values need to be compared to the generating values of 0.00 and 1.5, respectively.

3.6 Discussion

This chapter presented an introduction to the Rasch model and the estimation of its parameters using a customary method that capitalized on the parameter separability offered by the existence of simple sufficient statistics for this exponential family model. It was pointed out that, while there are alternative unbiased estimation methods, only CML estimation makes use of this property to the fullest extent. However, the conditional approach is only applicable to the Rasch model and a few of its extensions, while the LLCA approach presented above can also be used for more highly parameterized IRT models (e.g., Heinen, 1996; Haberman et al., 2008; von Davier, 2008).

The Rasch model is used operationally in international as well as national test and assessment programs. A variety of software programs for consistent estimation (e.g.,

Pfanzagl, 1994) of its parameters are available, including Conquest (Wu et al., 1997; Adams et al., Volume Three, Chapter 27), LPCM-WIN (Fischer and Ponocny-Seliger, 1998), Logimo (Kelderman and Steen, 1988), Multira (Carstensen and Rost, 2003), RSP (Glas and Ellis, 1994), RUMM2010 (Andrich, 2001), and WINMIRA (von Davier, 1994, 2001). Most of these software packages also offer model extensions such as the ability to estimate Rasch models for polytomous data (e.g., Volume One, Chapters 1 and 7), models with covariates of person parameters (Volume One, Chapters 32 and 33), and (dichotomous as well as polytomous) models for multiple populations (Volume One, Chapter 23). In addition, there exist add-ons to general-purpose statistical software such as R (R Core Team 2012), STATA, and SAS; for example, the eRm (Mair and Hatzinger, 2007) and ltm packages (Rizopoulos, 2006). These add-ons, due to their nature of being implemented as interpreted scripts in a general-purpose software package, are typically somewhat less suitable for very large datasets as they offer lower computational performance than dedicated programs.

The Rasch model will remain the model of choice for some researchers, while it will remain the "too restrictive" IRT model "that never fits the data" (de Leeuw, 1997) for others. While it is often found that a more general IRT model will fit the data better, Molenaar (1997) reminds us that even with some misfit, use of the Rasch model may still be warranted when the conclusions drawn from this model differ little from a higher parameterized unidimensional IRT model. This does not mean that the issue of model fit should be ignored. As with any other model, it is imperative to examine whether the Rasch model fits the data satisfactorily.

References

Andersen, E. B. 1972. The numerical solution to a set of conditional estimation equations. *Journal of the Royal Statistical Society, Series B*, 34, 283–301.

Andersen, E. B. 1973. A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123–140.

Andersen, E. B. 1977. Sufficient statistics and latent trait models. *Psychometrika*, 42, 69–81.

Andersen, E. B. 1980. *Discrete Statistical Models with Social Science Applications.* Amsterdam, The Netherlands: North-Holland Publishing Company.

Andersen, E. B., and Madsen, M. 1977. Estimating the parameters of the latent population. *Psychometrika*, 42, 357–374.

Andrich, D. 2001. *RUMM2010: Rasch Unidimensional Measurement Models*. Perth, Australia: RUMM Laboratory.

Bickel, P. J., and Doksum, K. A. 1977. *Mathematical Statistics: Basic Ideas and Selected Topics.* San Francisco, CA: Holden-Day, Incorporated.

Bradley, R. A., and Terry, M. E. 1952. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 39, 324–345.

Carstensen, C. H., and Rost, J. 2003. *Multira [Computer Software]*. Bamberg, Germany: University of Bamberg. Retrieved from http://www.multira.de/.

de Leeuw, J. 1997. Review of Rasch models—Foundations, recent developments, and applications. *Statistics in Medicine*, 16, 1431–1434.

de Leeuw, J., and Verhelst, N. 1986. Maximum-likelihood-estimation in generalized Rasch models. *Journal of Educational Statistics*, 11, 183–196.

Dennis, J. E., and Schnabel, R. B. 1996. Numerical methods for unconstrained optimization and nonlinear equations. *SIAM Classic in Applied Mathematics*, 16, 1996.

Eggen, T. J. H. M. 2000. On the loss of information in conditional maximum likelihood estimation of item parameters. *Psychometrika*, 65(3), 337–362.

Firth, D. 1993. Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27–38.

Fischer, G. H. 1981. On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, 46, 59–77.

Fischer, G. H. 2006. Rasch models. In C. R. Rao and S. Sinharay (Eds.), *Handbook of Statistics, Vol. 26: Psychometrics* (pp. 515–586). Amsterdam, The Netherlands: Elsevier Science Publishers.

Fischer, G. H., and Molenaar I. W. 1995. *Rasch Models: Foundations, Recent Developments and Applications*. New York, NY: Springer.

Fischer, G. H., and Ponocny-Seliger, E. 1998. *Structural Rasch Modeling. Handbook of the Usage of LPCM-WIN 1.0*. Groningen, The Netherlands: Institute ProGamma.

Follmann, D. 1988. Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika*, 53, 553–562.

Formann, A. K. 1994. Measuring change in latent subgroups using dichotomous data: Unconditional, conditional, and semiparametric maximum-likelihood-estimation. *Journal of the American Statistical Association*, 89, 1027–1034.

Gelman, A. 2007. http://andrewgelman.com/2007/05/11/weakly_informat/

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2003. *Bayesian Data Analysis* (2nd ed.). New York, NY: Chapman and Hall/CRC Texts in Statistical Science.

Glas, C. A. W. 1988a. The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika*, 53, 525–546.

Glas, C. A. W. 1988b. The Rasch model and multistage testing. *Journal of Educational Statistics*, 13, 45–52.

Glas, C. A. W. 2007. Testing generalized Rasch models. In M. von Davier and C. H. Carstensen (Eds.), *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications* (pp. 37–55). New York, NY: Springer.

Glas, C. A. W., and Ellis, J. L. 1994. *RSP: The Rasch Scaling Package [Computer Program]*. Groningen, The Netherlands: Institute ProGamma.

Glas, C. A. W., and Verhelst, N. D. 1995. Testing the Rasch model. In G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models: Their Foundations, Recent Developments and Applications* (pp. 69–96). New York, NY: Springer.

Gustafsson, J.-E. 1980. A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement*, 40, 327–385.

Haberman, S. J. 1977. Maximum likelihood estimates in exponential response models. *Annals of Statistics*, 5, 815–841.

Haberman, S. J. 2004. *Joint and Conditional Maximum Likelihood Estimation for the Rasch Model for Binary Responses* (Research Report No. RR-04-20). Princeton, NJ: Educational Testing Service.

Haberman, S. J., von Davier, M., and Lee, Y. 2008. *Comparison of Multidimensional Item Response Models: Multivariate Normal Ability Distributions versus Multivariate Polytomous Ability Distributions*. (Research Report No. RR-08-45). Princeton, NJ: Educational Testing Service.

Heinen, T. 1996. *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks, CA: Sage.

Kelderman, H. 1984. Loglinear Rasch model tests. *Psychometrika*, 49, 223–245.

Kelderman, H., and Steen, R. 1988. *LOGIMO: A Program for Loglinear IRT Modeling*. Enschede, The Netherlands: University of Twente, Department of Education.

Kiefer, J., and Wolfowitz, J. 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, 27, 887–906.

Klauer, K. C. 1991. An exact and optimal standardized person fit test for assessing consistency with the Rasch model. *Psychometrika*, 56(2), 213–228.

Lindsay, B., Clogg, C. C., and Grego, J. 1991. Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, 86, 96–107.

Mair, P., and Hatzinger, R. 2007. Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20, 1–20.

Martin-Löf, P. 1973. Statistiska Modeller. *Anteckningar fran seminarier läsaret 1969–70 utarbetade av Rolf Sundberg. Obetydligt ändrat nytryk*. Stockholm, Sweden: Institutet för försäkringsmatematik och matematisk statistik vid Stockholms universitet.

Molenaar, I. W. 1983. Some improved diagnostics for failure in the Rasch model. *Psychometrika*, 48, 49–72.

Molenaar, I. W. 1997. Lenient or strict application of IRT with an eye on practical consequences. In J. Rost and R. Langeheine (Eds.), *Applications of Latent Trait and Latent Class Models in the Social Sciences* (pp. 38–49). Münster, Germany: Waxmann.

Pfanzagl, J. 1994. On item parameter estimation in certain latent trait models. In G. H. Fischer and D. Laming (Eds.), *Contributions to Mathematical Psychology, Psychometrics, and Methodology* (pp. 249–263). New York, NY: Springer.

Ponocny, I. 2000. Exact person fit indexes for the Rasch model for arbitrary alternatives. *Psychometrika*, 65, 29–42.

R Core Team 2012. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: University of Chicago Press.

Rasch, G. 1966a. An individualistic approach to item analysis. In P. F. Lazarsfeld and N. W. Henry (Eds.), *Readings in Mathematical Social Science* (pp. 89–107). Chicago, IL: Science Research Associates.

Rasch, G. 1966b. An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49–57.

Rizopoulos, D. 2006. ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1–25.

Rost, J. 2004. *Lehrbuch Testtheorie—Testkonstruktion* [Textbook test theory—Test construction]. Bern, Germany: Huber.

Rost, J., and von Davier, M. 1994. A conditional item fit index for Rasch models. *Applied Psychological Measurement*, 18, 171–182.

Rost, J., and von Davier, M. 1995. Mixture distribution Rasch models. In G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments and Applications* (pp. 257–268). New York, NY: Springer.

Thissen, D. 1982. Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47, 175–186.

van den Wollenberg, A. L. 1982. Two new test statistics for the Rasch model. *Psychometrika*, 47, 123–139.

Verguts, T., and De Boeck, P. 2000. A note on the Martin-Löf test for unidimensionality. *Methods of Psychological Research Online*, 5, 77–82. Retrieved from http://ppw.kuleuven.be/okp/_pdf/Verguts2000ANOTM.pdf

Verhelst, N. 2001. Testing the unidimensionality assumption of the Rasch model. *Methods of Psychological Research Online 2001*, 6(3), 231–271. Retrieved from http://136.199.86.12/fachgruppen/methoden/mpr-online/issue15/art2/verhelst.pdf

Verhelst, N. D., and Glas, C. A. W. 1995. The generalized one-parameter model: OPLM. In G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments and Applications* (pp. 215–238). New York, NY: Springer.

von Davier, M. 1994. *WINMIRA—A Windows Program for Analyses with the Rasch Model, with the Latent Class Analysis and with the Mixed Rasch Model*. Kiel, Germany: Institute for Science Education.

von Davier, M. 2001. *WINMIRA 2001: Software for Estimating Rasch Models, Mixed and Hybrid Rasch Models, and the Latent Class Analysis*. Manual retrieved from http://www.von-davier.com

von Davier, M. 2008. A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287–307.

von Davier, M., and Carstensen. C. H. (Eds.) 2007. *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*. New York, NY: Springer.

von Davier, M., and Molenaar, I. W. 2003. A person-fit index for polytomous Rasch models, latent class models, and their mixture generalizations. *Psychometrika*, 68, 213–228.

von Davier, M., and Rost, J. 1995 Polytomous mixed Rasch models. In G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments and Applications* (pp. 371–379). New York, NY: Springer.

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

Wright, B. D., and Linacre, J. M. 1991. *BIGSTEPS Computer Program for Rasch Measurement*. Chicago, IL: MESA Press.

Wright, B. D., and Stone, M. H. 1979. *Best Test Design: Rasch Measurement*. Chicago, IL: MESA Press.

Wu, M. L., Adams, R. J., and Wilson, M. R. 1997. *ConQuest: Multi-Aspect Test Software*. Camberwell, Australia: Australian Council for Educational Research.

Xu, X., and Jia, Y. 2011. *The Sensitivity of Parameter Estimates to the Latent Ability Distribution* (Research Report No. RR-11-40). Princeton, NJ: Educational Testing Service.

Xu, X., and von Davier, M. 2008. *Comparing Multiple-Group Multinomial Loglinear Models for Multidimensional Skill Distributions in the General Diagnostic Model* (Research Report No. RR-08-35). Princeton, NJ: Educational Testing Service.

Zermelo, E. 1929. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung [The calculation of tournament results as a maximum problem of probability calculus]. *Mathematische Zeitschrift*, 29, 436–460.

Zwinderman, A. H. 1995. Pairwise estimation in the Rasch models. *Applied Psychological Measurement*, 19, 369–375.