

Article

A Comprehensive Simulation Study of Estimation Methods for the Rasch Model

Alexander Robitzsch ^{1,2} 
¹ IPN—Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, 24118 Kiel, Germany; robitzsch@leibniz-ipn.de

² Centre for International Student Assessment (ZIB), Olshausenstraße 62, 24118 Kiel, Germany

Abstract: The Rasch model is one of the most prominent item response models. In this article, different item parameter estimation methods for the Rasch model are systematically compared through a comprehensive simulation study: Different alternatives of joint maximum likelihood (JML) estimation, different alternatives of marginal maximum likelihood (MML) estimation, conditional maximum likelihood (CML) estimation, and several limited information methods (LIM). The type of ability distribution (i.e., nonnormality), the number of items, sample size, and the distribution of item difficulties were systematically varied. Across different simulation conditions, MML methods with flexible distributional specifications can be at least as efficient as CML. Moreover, in many situations (i.e., for long tests), penalized JML and JML with ϵ adjustment resulted in very efficient estimates and might be considered alternatives to JML implementations currently used in statistical software. Moreover, minimum chi-square (MINCHI) estimation was the best-performing LIM method. These findings demonstrate that JML estimation and LIM can still prove helpful in applied research.

Keywords: Rasch model; estimation methods; nonnormality



Citation: Robitzsch, A.

A Comprehensive Simulation Study of Estimation Methods for the Rasch Model. *Stats* **2021**, *4*, 814–23.

<https://doi.org/10.3390/stats4040048>

Academic Editor: Wei Zhu

Received: 31 August 2021

Accepted: 28 September 2021

Published: 1 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Rasch model (RM; [1–3]) is one of the most popular item response theory (IRT) models [4–9]. It is important to select appropriate estimation methods because the RM is widespread in diverse applications (e.g., [10–16]). For the RM, a variety of estimation methods has been proposed. In this article, a comprehensive comparison of different estimation methods for the RM is conducted. We manipulate the factors' test length (i.e., number of items), sample size, the type of ability distribution, and the distribution of item difficulties. Recommendations for the choice of estimation methods can be drawn for empirical applications that utilize the RM.

The article is structured as follows. In Section 2, the RM model is introduced. In Section 3, several estimation methods are reviewed. In Section 4, we present the results of a simulation study that compares a wide range of estimation methods. Finally, the paper closes with a discussion in Section 5.

2. Rasch Model

The RM [1,2,17–27] is a statistical model for dichotomous item responses X_{pi} for persons $p = 1, \dots, N$ and items $i = 1, \dots, I$. It assumes the existence of a latent variable θ (so-called ability) that accounts for the dependence among item responses. The item response function for the Rasch model is given as

$$P(X_{pi} = x | \theta_p; b_i) = \frac{\exp(x(\theta_p - b_i))}{1 + \exp(\theta_p - b_i)}, \quad x = 0, 1, \quad (1)$$

where θ_p is the ability of person p and b_i is the item difficulty for item i . Abilities θ_p can be either modeled as fixed effects or random effects [28,29]. In the treatment of fixed

effects, every person is associated with an ability parameter that has to be estimated. In the random effects situation, a distribution for the ability variable is posed; that is, $\theta \sim G$ and the unknown distribution G must be estimated in a parametric, semiparametric, or nonparametric way. Note that the RM parameters are determined up to a constant. Hence, either the mean of the abilities or the mean of item difficulties has to be fixed to zero for reasons of identification [30]. The posed functional form of the item response function (1) in the RM can be assessed by item fit statistics [31]. We would also like to emphasize that the RM only places low requirements on sample sizes because only one parameter per item (i.e., item difficulty b_i) is estimated [9].

In addition to Equation (1), item responses X_{pi} are assumed to be locally independent:

$$P(X_{p1} = x_1, \dots, X_{pI} = x_I | \theta_p) = \prod_{i=1}^I P(X_{pi} = x_i | \theta_p) \quad . \quad (2)$$

This means that there does not exist residual associations among items after taking the ability θ_p into account. Assumption (2) can be tested in empirical applications [32–34]. It can be argued that the unidimensionality assumption in Equation (2) is only a crude approximation to real data and enables the extraction of a summary ability variable. The local independence assumption can then be understood as an assumption that residual associations among items cancel out on average. This means that there will always exist positive and negative residual associations after controlling for the extracted ability variable θ .

An important property of the Rasch model is that the sum score $S_p = \sum_{i=1}^I X_{pi}$ is a sufficient statistic for θ_p [30]. Hence, θ_p is a nonlinear function of the sum score S_p , and it does not matter in the computation of the ability which of the items have been solved by a person. Moreover, with at least a moderate number of items, the nonlinear relation of S_p and θ_p can be closely approximated by a linear function which explains the resemblance of classical test theory [35] and the RM [36]. Moreover, note that the proportion correct for an item is a sufficient statistic for the item difficulty b_i .

In this article, the RM is a mixed effects logistic model with a random person effect θ , and item difficulties b_i are fixed effects [28,37–41]. The formulation of the RM as a mixed effects model has the advantage that item difficulties can alternatively be considered as random effects [28]. Moreover, more complex hierarchical structures (e.g., students nested within schools) can also be accommodated [39,42].

In the remainder of the paper, we only focus on the estimation of item parameters. We review several estimation methods for the RM in the next section.

3. Estimation Methods for the Rasch Model

A variety of estimation methods has been proposed for the RM [23,43,44]. In this section, we contrast joint maximum likelihood, conditional maximum likelihood, marginal maximum likelihood estimation, and limit information estimation methods.

3.1. Joint Maximum Likelihood (JML) Estimation

Joint maximum likelihood (JML; [25,45]) methods treat person abilities θ_p as fixed effects. In JML, the vector of person parameters $\gamma = (\theta_1, \dots, \theta_N)$ is simultaneously estimated with the vector of item parameters $\mathbf{b} = (b_1, \dots, b_I)$. The estimation JML algorithm alternates between γ and $\mathbf{b}$ parameter estimation in one iteration. Note that the number of estimated parameters grows with the number of observations (i.e., number of persons times number of items). This property is also denoted as the incidental parameter problem, resulting in the undesirable property that JML estimates are not consistent [46–48]. However, several bias correction methods can be utilized to circumvent this issue. The different JML estimation variants are described in more detail in the following.

3.1.1. JML with Bias Correction (JMLM and JMLW)

As mentioned above, the JML estimation algorithm cycles between the steps of estimating person and item parameters. For persons that solved no or all items, no finite ability estimate θ_p exist, which causes the incidental parameter problem. The JMLM method eliminates the persons with these extreme scores from the JML estimation for estimating item parameters. In contrast, a modified ability estimation method by Warm [49] can be used (JMLW) that results in finite ability estimates and does not require the elimination of persons in the analysis. Interestingly, the JMLW method can be interpreted as a Bayesian estimation method with a Jeffrey's prior for abilities [50]. The bias due to incidental parameters can be corrected (or at least reduced) in JMLM and JMLW by a subsequent adjustment of estimated item parameters [51,52]. With obtained item parameters $\hat{b}_i$ from the alternating estimating approach, the finally computed bias-corrected item parameter is given as $(I - 1)/I \cdot \hat{b}_i$. Note that the adjustment becomes negligible with an increasing number of items I .

3.1.2. Penalized JML (PJML)

In penalized JML [53–55], a ridge penalty term is added to the log-likelihood function with a regularization parameter λ . That is, a term $\text{Pen}(\theta_p) = -\lambda\theta_p^2$ is added to the person-specific log-likelihood. Including a ridge penalty is equivalent to a Bayesian approach in which a normal prior distribution $\theta \sim N(0, \sigma_{\text{prior}}^2)$ with an appropriate choice of the regularization parameter $\sigma_{\text{prior}} > 0$ is employed. PJML also circumvents the exclusion of persons with extreme scores from the estimation. The optimal choice of σ_{prior} will typically differ in the situations in which the precision in person or item parameter estimates should be optimized.

3.1.3. JML with ε Adjustment (JML ε)

Another JML estimation approach that does not require eliminating persons with extreme scores is JML with ε adjustment (JML ε ; [56–58]). JML ε estimation employs a modified likelihood by replacing the sufficient statistic S_p with a modified sufficient statistic S_p^* that is defined by

$$S_p^* = \varepsilon + \frac{I - 2\varepsilon}{I} \cdot S_p \quad , \quad (3)$$

using an appropriate $\varepsilon > 0$. As a consequence, while S_p takes values in the interval $[0, I]$, S_p^* takes values in $[\varepsilon, I - \varepsilon]$, and the latter statistic results in finite ability estimates.

Interestingly, the estimation methods PJML and JML ε tackle the issue of non-finite ability estimates from different angles. The original JML approach (i.e., JMLM, that does not allow persons with extreme scores) seeks ability estimates θ_p that solves the estimating equation

$$S_p = f(\theta_p) \quad . \quad (4)$$

The PJML method adds a penalty $\text{Pen}(\theta_p)$ to the right side of Equation (4); that is, $S_p = f(\theta_p) + \text{Pen}(\theta_p)$. The JML ε method changes the left side of Equation (4), resulting in the modified estimating equation $S_p^* = f(\theta_p)$.

3.2. Conditional Maximum Likelihood (CML) Estimation

Conditional maximum likelihood (CML; [43,59–62]) estimation can handle the situations in which the ability variable is either treated as fixed or random. In CML estimation, the vector of item difficulties $\mathbf{b}$ is only estimated. The ability variable θ is removed from the estimation by conditioning on the sum score. One can show that the conditional distribution of item responses $\mathbf{X}_p$ conditioned on the sum score $S_p = \sum_{i=1}^I X_{pi}$ does not depend on θ_p [30]:

$$P(\mathbf{X}_p = \mathbf{x}_p | S_p = s_p) = h(\mathbf{b}) \quad . \quad (5)$$

Hence, no distributional assumption about the ability variable has to be posed. In addition, item parameter estimates are consistent. CML estimation is computationally more demanding than JML, but efficient algorithms have been proposed [63,64].

CML estimation has also been discussed for mixed effects logistic models [65–67].

3.3. Marginal Maximum Likelihood (MML) Estimation

In marginal maximum likelihood estimation (MML; [68,69]), latent variables θ are integrated out by posing a distributional assumption G_γ for θ , where distribution parameters γ are simultaneously estimated with $\mathbf{b}$. The log-likelihood function $l(\mathbf{b}, \gamma)$ is maximized. The log-likelihood contribution for person p is given by

$$l_p(\mathbf{b}, \gamma) = \log \left[\int \prod_{i=1}^I P(X_{pi} = x_{pi} | \theta; b_i) dG_\gamma(\theta) \right]. \quad (6)$$

If the parametric specification G_γ differs from the data-generating distribution H , biased item parameters can occur.

In the following subsections, different distributional specifications in MML are discussed. These MML variants differ in how deviations from normally distributed abilities are handled (see [70,71]).

3.3.1. MML with Normality Assumption (MMLN)

In most applications and the default of most IRT software packages [72,73], a normal distribution for θ is posed (MMLN). For identification of the parameters in the RM, the mean is set to zero, and the standard deviation σ is estimated along with the item parameters $\mathbf{b}$. The integral in the log-likelihood function (6) is evaluated by numerical integration techniques. The consequences of applying a misspecified normal distribution have been frequently studied in the literature [74–77].

Different numerical approximations of the unidimensional integral involved in the likelihood function (see Equation (6)) have been proposed in the literature [78–80]. In our experience, numerical approximations defined as the default in IRT packages such as *mirt* [72] occasionally provide more accurate than corresponding defaults in the popular mixed effects R package *lme4* [37].

3.3.2. MML with Multinomial Distribution (MMLMN)

MML with a multinomial distribution (MMLMN) estimates a discrete distribution for the ability variable θ (see [81]). A fixed grid of θ points $\theta_1, \dots, \theta_C$ is chosen (e.g., on a grid of equidistant θ points ranging from -4 to 4 , see [82]). In MMLMN, probabilities $\gamma_c = P(\theta = \theta_c)$ are freely estimated. The number of estimated parameters increases with larger number C of grid points. An appropriate number C of discrete grid points must be chosen to ensure sufficiently stable item and distribution parameters estimation.

3.3.3. MML with Log-Linear Smoothing (MMLLS)

MML with log-linear smoothing (MMLLS) avoids estimating a large number of distribution parameters in MMLMN. In this estimation method, a log-linear smoothing on the discrete probabilities $\gamma_c = P(\theta = \theta_c)$ is performed [82,83] (see also [84,85]). If the first two moments are smoothed, MMLLS corresponds to the estimation of discretized normal distribution. In empirical applications, smoothing is typically performed up to the first three or four moments [86–88]. These higher moments capture deviations from normality. The log-linear smoothing approach can also be extended to handle nonlinear relations among several latent variables [86].

3.3.4. MML with Located Latent Classes (MMLLC)

The estimation methods MMLMN and MMLLS presuppose the specification of the discrete grid of θ points. In MML with located latent classes, for C latent classes, the values of the grid points θ_c are estimated in addition to probabilities γ_c (MMLLLC; [89–92]). In

the RM with a test of I items, at most $I/2$ located class models can be specified because model parameters in larger models cannot be identified [89]. Notably, MMLLC poses the weakest assumptions about the data-generating distribution G , but it relies on a possible doubtful discrete representation of the θ distribution. Classifying persons into different discrete ability levels might be conceptually appealing in empirical applications [93,94].

3.4. Limited Information Estimation Methods

So-called limited information methods (LIM) for estimating item parameters in the RM do not rely on the full item response pattern $\mathbf{x}_p$. These methods are often simpler to compute because they do not iterate through all item response patterns. LIM only consider marginal univariate or bivariate frequency distributions of item responses.

3.4.1. Pairwise Marginal Maximum Likelihood (PMML)

Pairwise MML (PMML; [95–98]) is a composite likelihood estimation method for which only pairwise item response probabilities $P(X_{pi} = x_{pi}, X_{pj} = x_{pj})$ are modelled. The contributions of all item pairs (i, j) are taken into account. In principle, any distributional assumption about θ can be posed, like in joint modeling of the probabilities as in MML. However, a normal distribution is often assumed [95,99].

3.4.2. Pairwise Conditional Maximum Likelihood (PCML)

In pairwise CML (PCML; [100–104]), the conditional probabilities $P(X_{pi} = x_{pi}, X_{pj} = x_{pj}) / P(X_{pi} + X_{pj} = x_{pi} + x_{pj})$ are used for defining an optimization function. Like for CML that conditions on the sum score, PCML also removes θ from estimation equations and does not pose distributional assumptions. The advantage of PCML compared to CML is the strongly reduced burden in computational demand.

3.4.3. Minimum Chi-Square Method (MINCHI)

Minimum chi-square (MINCHI) estimation only relies on bivariate frequencies f_{ij} that are defined as

$$f_{ij} = P(X_{pi} = 1, X_{pj} = 0) \quad . \quad (7)$$

In MINCHI, the following squared distance is defined that is minimized for determining item parameter estimates $\mathbf{b}$ (see [30,105,106]):

$$h(\mathbf{b}) = \sum_{i,j} \frac{(\epsilon_j^{-1} f_{ij} - \epsilon_i^{-1} f_{ji})^2}{\epsilon_i^{-1} \epsilon_j^{-1} (f_{ij} + f_{ji})} \quad , \quad (8)$$

where $\epsilon_i = \exp(-b_i)$. Fixed-point estimation equations have been proposed for computing the minimizer of Equation (8) (see [30]). Also, note that no distributional assumptions about θ are required for MINCHI estimation.

3.4.4. Row Averaging Method (RA)

Like MINCHI estimation, the row averaging method [107–109] relies on bivariate frequencies f_{ij} (see Equation (7)). A matrix $\mathbf{B}$ with entries $b_{ij} = \log(f_{ij}/f_{ji})$ is formed. The row-wise average of entries in the matrix $\mathbf{B}$ is used as an item parameter estimate [107,110]. If some cells (i, j) are empty, $\mathbf{B}$ cannot be computed. An alternative estimation method involving powers has been proposed. Let $\mathbf{F}$ denote the matrix consisting of all elements f_{ij} (the so-called incidence matrix). The computation of $\mathbf{B}$ can then rely on entries of the matrix $\mathbf{F}^* = \mathbf{F}^k$, where k is an integer larger than one (e.g., 2 or 3).

3.4.5. Eigenvector Method (EVM)

The eigenvector method (EVM; [111–114]) relies on the same preprocessing steps as RA. However, instead of row averaging, the first eigenvector of $\mathbf{B}$ is computed as the estimate of the vector of item difficulties. In the case of empty cells, power matrices $\mathbf{F}^* = \mathbf{F}^k$

(see Section 3.4.4) can be used. Note that RA and EVM do not require iterations and only require low computational demands that might be attractive for large-scale applications.

3.4.6. Log-Linear by Linear Association Models (LLLA)

Log-linear by linear association (LLLA; [115–118]) models estimate item parameters through a pseudo-likelihood approach. It relies on the fact that a logistic regression for $P(X_{pi} = 1|S_p - X_{pi})$ of the item response X_{pi} on the rest score $S_p - X_{pi}$ can be specified in which θ does not appear (assuming a normal distribution of θ ; see [119]). The logistic regression stacks data of all item responses and allows simultaneous estimation of all item parameters [118,120].

4. Simulation Study

4.1. Purpose

Many simulation studies compare the performance of different item parameter estimation methods for the RM. However, most studies only considered the main estimation methods CML, JML, and MML (see, e.g., [44,77,110,121] for the RM and [122] for the mixed effects logistic model). Moreover, they often only used limited variations of deviations from normality in the ability variable. In this study, we provide a comprehensive comparative simulation study that compares the performance of a large number of estimation methods under a wide range of θ distribution. Moreover, sample size, the number of items, and the distribution of item difficulties are systematically manipulated. This simulation study systematically extends the simulation design employed in [123].

4.2. Design

In the simulation study, item response data has been generated for the RM. We manipulated six factors in the simulation. First, the sample size (N) of persons was manipulated, resulting in three levels $N = 250, 500, 1000$. These sample sizes reflect small-scale to large-scale applications of the RM. Second, we varied the number of items (I) and chose levels $I = 10$ and $I = 30$. In the simulation, a set of item difficulties was specified for the $I = 10$ condition. In the condition $I = 30$, the item parameters for $I = 10$ were used three times. The levels reflect a short and a long test in applications. Third, the range of item difficulties was manipulated. In the condition of a test with a symmetric item difficulty distribution, item parameters were chosen from the interval $[-3, 3]$ for a wide range, and from the interval $[-1.5, 1.5]$ for a small range. Fourth, the skewness of item difficulties was varied. In the symmetric case, item parameters were equidistantly chosen from the intervals $[-3, 3]$ and $[-1.5, 1.5]$, respectively. In the case of a skew item difficulty distribution, larger item difficulties appear more frequently than smaller item difficulties. The precisely chosen item parameters can be found in Appendix A. Fifth, eight data-generating distributions for the latent ability variable θ in the RM were simulated. All distributions were standardized, that is, $E(\theta) = 0$ and $SD(\theta) = 1$. The eight simulated θ distributions are:

1. NO: A normal distribution ($N(0, 1)$) with zero mean and a standard deviation of one
2. Chi²: A scaled chi-squared distribution with one degree of freedom
3. UN: A uniform distribution on the interval $[-1.73, 1.73]$ (i.e., $U(-1.73, 1.73)$)
4. BE: A scaled U-shaped beta distribution with shape and scale parameters of 0.5; that is, $\theta \sim 2.83 \cdot (\text{Beta}(0.5, 0.5) - 0.5)$
5. SM: A symmetric mixture distribution with $\theta = 0.898 \cdot \theta^*$, and $\theta^* \sim 0.5 \cdot N(-0.8, 0.77^2) + 0.5 \cdot N(0.8, 0.77^2)$
6. AM: An asymmetric mixture distribution with $\theta = 0.994 \cdot (\theta^* - 0.479)$, and $\theta^* \sim 0.2 \cdot N(-0.8, 0.77^2) + 0.8 \cdot N(0.8, 0.77^2)$
7. LC2: A discrete distribution with θ points $-2.0, 0.5$ and corresponding probabilities 0.20 and 0.80

8. LC3: A discrete distribution with θ points $-0.790, 1.033, 2.248$ and corresponding probabilities $0.60, 0.35$, and 0.05

In total, there were $3 \times 2 \times 2 \times 2 \times 8 = 192$ conditions employed in the simulation. In total, 1000 datasets were simulated and analyzed in each condition.

4.3. Analysis Models

Item parameters for the simulated datasets were estimated with the different methods discussed in Section 3. Throughout the simulation, we only considered the estimation of item parameters and did not consider person parameter estimation. To enable the comparability of item parameter estimates, we centered estimated item parameters obtained from each estimation method (i.e., they have zero mean).

For the PJML estimation (see Section 3.1.2), we chose normal priors $N(0, \sigma_{\text{prior}}^2)$ with $\sigma_{\text{prior}} = 1, 1.5$, and 2 . Notably, an optimal value of σ_{prior} could also be estimated by cross-validation or empirical Bayes methods. For JML ϵ estimation (see Section 3.1.3), we specified values $\epsilon = 0.1, 0.2, 0.24, 0.3, 0.4, 0.5$. The value $\epsilon = 0.24$ turned out to be optimal in preliminary simulation studies. For MMLMN estimation (see Section 3.3.2), we specified models with 5 equidistant θ grid points in $[-2, 2]$, 7 equidistant θ grid points in $[-3, 3]$, 11 equidistant θ grid points in $[-4, 4]$, and 15 equidistant θ grid points in $[-4, 4]$. For MMLLS estimation (see Section 3.3.3), we used log-linear smoothing up to three and four moments. The inclusion of moments larger than two allows deviations from normality. An equidistant grid of 15 ability values in $[-4, 4]$ was chosen. For MMLLC (see Section 3.3.4), we specified analysis with 2, 3, 4, and 5 located latent classes. Notably, the data-generating models LC2 and LC3 are expected to be properly handled by one of these models. For RA estimation (see Section 3.4.4), we used powers 1, 2, and 3 of the incidence matrix F for the estimation (i.e., use $F^* = F^k$ with $k = 1, 2, 3$ as the basis for the computation of the matrix B). For EVM estimation (see Section 3.4.5), powers 2 and 3 of the incidence matrix F were utilized.

The whole simulation was carried out in the statistical software R [124] utilizing the R packages immer [125] (CML, JML ϵ), pairwise [126] (LLLA), plRasch [118,127] (RA) and sirt [128] (EVM, JMLM, JMLW, MINCHI, MMLLC, MMLLS, MMLMN, MMLN, PCML, PJML). For PMML, a dedicated function was implemented in R.

4.4. Outcome Measures

The bias and root mean square error (RMSE) was computed for each estimated item parameter $\hat{b}_i$. We consider two summary measures of item parameter recovery. First, the mean absolute bias MAB (also labeled as bias in the Results Section 4.5)

$$\text{MAB}(\hat{\mathbf{b}}) = \frac{1}{I} \sum_{i=1}^I |\text{Bias}(\hat{b}_i)| \quad (9)$$

quantifies the average bias of item parameters. MAB values near to zero indicate situations with unbiased item parameter estimates.

Second, bias and variability are summarized in the average RMSE (ARMSE) defined by

$$\text{ARMSE}(\hat{\mathbf{b}}) = \frac{1}{I} \left[\sum_{i=1}^I \text{RMSE}(\hat{b}_i) \right] \quad (10)$$

To ease the comparison of different estimation methods independent of sample size, ARMSE values are normed with respect to the best-performing estimation method (with a corresponding value $\text{ARMSE}_{\text{best}}(\hat{\mathbf{b}})$ in one replication cell in the simulation. The so-called relative RMSE (RRMSE) is defined as

$$\text{RRMSE}(\hat{\mathbf{b}}) = \frac{\text{ARMSE}(\hat{\mathbf{b}})}{\text{ARMSE}_{\text{best}}(\hat{\mathbf{b}})} \quad (11)$$

As a consequence, RRMSE values have an optimal value of 100, which are attained by the best-performing estimation method.

To summarize the contribution of each of the manipulated factors in the simulation, we conducted an analysis of variance (ANOVA) based on a linear regression model and used a variance decomposition for assessing the importance (i.e., computed the eta square effect size).

Moreover, we classified estimation methods whether they showed acceptable performance in a particular condition. We defined acceptable performance for the bias if the bias (i.e., the MAB) was smaller than 0.025. Assuming a symmetric item difficulty distribution and that bias is proportional to the true item difficulty, this condition would correspond to a maximum item parameter bias of 0.05. An estimator had satisfactory performance concerning the RMSE if the relative RMSE was smaller than 107. This is equivalent to an average loss in precision in estimated item parameters of about 15% (i.e., $1.07^2 = 1.144$).

4.5. Results

In Table 1, the variance decomposition from the ANOVA of different simulation factors in the simulation for bias and the (relative) RMSE is presented. All terms up to three-way interactions were included. From the size of the residual variance, it can be concluded that the first three orders capture the most important sources of variance of simulation factors. It turned out that the estimation method (Meth) was the most important first-order factor for bias and RMSE, followed by the range of item difficulties (Range) and the number of items (I). The performance of estimation methods for bias and RMSE depended on an interaction effect with the number of items. Interestingly, there was only an interaction effect of estimation and sample size (N) for the RMSE and not for the bias. Moreover, the performance of estimation methods was also moderated by the range of the item difficulty distribution. Finally, there were also some important three-order terms involving estimation methods (i.e., $N \times I \times \text{Meth}$, $N \times \text{Range} \times \text{Meth}$, $I \times \text{Range} \times \text{Meth}$). For a selected number of cells, we present some results that demonstrate these interaction effects in more detail.

In Table 2, the performance of the different estimation methods for bias and RMSE are summarized across 192 conditions of the simulation. CML and the LIM MINCHI (being the best estimation method with respect to bias), PCML, EVM, and RA (with powers of the incidence matrix larger than 1) are approximately unbiased across simulation conditions. For MML estimation methods, only those methods were unbiased that specified the ability distribution flexible enough. For the multinomial modeling (MMLMN), a large number of θ grid points (11 or 15; MMLMN(11) or MMLMN(15)) was needed for producing acceptable performance in most of the simulation conditions. At least three located latent classes (MMLLC) were needed for an acceptable estimation of item parameters with respect to bias. Notably, estimation under the normal distribution (MMLN) or with only two latent classes (MMLLC(2)) was unsuccessful in a variety of conditions. Furthermore, all JML variants showed biased item parameter estimates. PJML with a prior of $\sigma_{\text{prior}} = 1.5$ turned out to be the best-performing method with respect to bias throughout all simulation conditions. Interestingly, the method JMLM that eliminates persons from estimation resulted in a smaller bias than JMLW that does not remove persons.

For the RMSE, JML ϵ with $\epsilon = 0.24$ performed best. Like for bias, only sufficiently flexible distributional specifications in MML resulted in acceptable performance for the RMSE. It was indicated to use four instead of only three moments for log-linear smoothing (MMLLS). Again, located latent class models (MMLLC) produced relatively precise item parameter estimates that were even superior to estimates obtained from CML. LIM showed higher RMSE values compared to MML variants and CML. However, the best-performing LIM MINCHI outperformed MML with normal distribution assumption (MMLN) and the widely implemented JML variants JMLM and JMLW. In particular, MINCHI (and partly PCML) should be preferred over EVM and RA estimation.

Table 1. Variance proportions of different factors in the simulation study for (mean absolute) bias and relative RMSE.

Source	Bias	Rel. RMSE
<i>N</i>	0.2	0.9
<i>I</i>	5.3	5.5
Skew	0.2	0.1
Range	6.3	6.5
Meth	51.6	29.3
Dist	0.4	0.1
<i>N</i> × <i>I</i>	0.0	0.6
<i>N</i> × Skew	0.0	0.0
<i>N</i> × Range	0.0	0.4
<i>N</i> × Meth	0.7	9.2
<i>N</i> × Dist	0.0	0.1
<i>I</i> × Skew	0.0	0.1
<i>I</i> × Range	0.7	1.4
<i>I</i> × Meth	17.5	21.3
<i>I</i> × Dist	0.0	0.0
Skew × Range	0.4	0.4
Skew × Meth	0.4	0.6
Skew × Dist	0.0	0.0
Range × Meth	7.9	7.3
Range × Dist	0.1	0.1
Meth × Dist	1.6	0.5
<i>N</i> × <i>I</i> × Skew	0.0	0.0
<i>N</i> × <i>I</i> × Range	0.0	0.0
<i>N</i> × <i>I</i> × Meth	0.0	4.4
<i>N</i> × <i>I</i> × Dist	0.0	0.0
<i>N</i> × Skew × Range	0.0	0.0
<i>N</i> × Skew × Meth	0.1	0.3
<i>N</i> × Skew × Dist	0.0	0.0
<i>N</i> × Range × Meth	0.1	1.4
<i>N</i> × Range × Dist	0.0	0.0
<i>N</i> × Meth × Dist	0.1	0.2
<i>I</i> × Skew × Range	0.1	0.1
<i>I</i> × Skew × Meth	0.2	0.3
<i>I</i> × Skew × Dist	0.0	0.0
<i>I</i> × Range × Meth	2.7	4.5
<i>I</i> × Range × Dist	0.0	0.0
<i>I</i> × Meth × Dist	0.2	0.1
Skew × Range × Meth	1.0	0.7
Skew × Range × Dist	0.0	0.0
Skew × Meth × Dist	0.1	0.1
Range × Meth × Dist	1.1	0.7
Residual	1.0	3.0

Note. *N* = sample size; *I* = number of items; Dist = simulated trait distribution; Meth = estimation method; Skew = Skewness of item difficulties; Range = Range in item difficulties. Percentage values larger than 0.5 are printed in bold.

Table 2. Summary of results for (mean absolute) bias and relative RMSE for different estimation methods across all simulation conditions.

Method	Bias					Relative RMSE				
	Rk	%Acc	Med	Q90	MAD	Rk	%Acc	Med	Q90	MAD
MMLN	17	81.3	0.013	0.040	0.009	14	74.5	104.2	109.9	3.1
MMLLS(3)	14	90.1	0.008	0.023	0.007	11	81.3	103.5	109.1	3.0
MMLLS(4)	3	100.0	0.006	0.014	0.003	5	93.8	102.4	106.2	2.5
MMLMN(5)	21	62.0	0.016	0.080	0.015	18	68.8	104.4	125.4	4.2
MMLMN(7)	15	85.9	0.009	0.029	0.007	10	81.8	103.4	108.4	3.2
MMLMN(11)	12	96.9	0.006	0.016	0.004	6	91.7	102.5	106.3	2.5
MMLMN(15)	9	99.5	0.006	0.016	0.004	7	91.7	102.6	106.7	2.5
MMLLC(2)	26	40.1	0.028	0.061	0.020	17	72.4	104.2	116.5	4.0
MMLLC(3)	13	94.3	0.009	0.022	0.006	3	94.8	102.6	105.7	2.5
MMLLC(4)	6	100.0	0.007	0.015	0.004	2	94.8	102.1	106.0	2.2
MMLLC(5)	4	100.0	0.007	0.014	0.004	4	94.3	102.3	106.3	2.3
CML	2	100.0	0.006	0.015	0.004	8	91.1	103.0	106.8	2.3
JMLM	23	54.2	0.021	0.132	0.022	21	56.8	105.9	145.6	5.2
JMLW	27	35.4	0.035	0.077	0.025	22	53.1	106.4	125.2	6.8
PJML(1.0)	30	22.4	0.048	0.111	0.031	23	45.8	108.1	134.2	9.1
PJML(1.5)	16	84.9	0.010	0.032	0.007	9	85.9	103.4	107.6	3.0
PJML(2.0)	28	27.1	0.038	0.085	0.024	30	26.0	111.1	129.9	6.7
JML ϵ (0.1)	29	25.0	0.053	0.174	0.036	32	23.4	116.0	179.4	13.9
JML ϵ (0.2)	24	53.6	0.024	0.052	0.017	13	79.2	103.4	109.5	2.5
JML ϵ (0.24)	22	56.3	0.023	0.040	0.015	1	95.3	101.1	104.8	1.6
JML ϵ (0.3)	25	46.4	0.036	0.077	0.030	16	72.4	101.3	119.4	1.9
JML ϵ (0.4)	31	3.1	0.065	0.166	0.050	28	41.7	109.6	161.1	14.3
JML ϵ (0.5)	32	0.0	0.101	0.248	0.069	31	25.0	125.5	215.3	33.3
PMML	20	70.3	0.015	0.067	0.011	20	59.9	105.7	120.9	4.5
PCML	5	100.0	0.007	0.017	0.004	19	62.0	106.1	111.0	2.9
LLLA	19	80.7	0.013	0.042	0.007	15	74.5	104.2	110.4	2.8
MINCHI	1	100.0	0.005	0.012	0.003	12	80.7	104.9	108.6	2.2
EVM(2)	11	98.4	0.007	0.019	0.004	25	41.7	108.9	117.8	6.6
EVM(3)	8	100.0	0.007	0.019	0.004	27	41.7	108.8	118.1	6.6
RA(1)	18	81.3	0.019	0.028	0.010	29	34.4	110.0	123.0	6.1
RA(2)	10	99.5	0.007	0.019	0.004	24	41.7	108.9	117.8	6.6
RA(3)	7	100.0	0.007	0.019	0.004	26	41.7	108.8	118.1	6.6

Note. CML = conditional maximum likelihood; EVM = eigenvector method; JML = joint maximum likelihood; JML ϵ = JML with ϵ adjustment; JMLM = JML with maximum likelihood ability estimator; JMLW = JML with Warm's maximum likelihood ability estimator; LLLA = log-linear by linear association method; MINCHI = minimum chi-square estimation; MML = marginal maximum likelihood; MMLLLC = MML with located latent classes; MMLLS = MML with log-linear smoothing; MMLMN = MML with multinomial distribution; MMLN = MML with normal distribution; PJML = penalized JML; PMML = pairwise MML; PCML = pairwise CML; RA = row-averaging method; Rk = performance rank of the method; %Acc = percentage of conditions with acceptable performance; Med = median; Q90 = 90% quantile; %Acc values larger than 90, biases smaller than 0.025 and RMSE values smaller than 107 are printed in bold.

Table 3 shows the bias and the RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 10$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution. Six out of eight data-generating models are depicted that demonstrate the most important differences among estimation methods. The MML method that poses a normal distribution assumption (i.e., MMLN) provides the least bias if the latent ability was generated with a normal distribution. The largest bias was obtained if the located latent class model (LC(2) or LC(3)) generated the data. If θ was normally distributed, log-linear smoothing (MMLLS) and a multinomial distribution (MMLMN) with at least 7 grid points provided approximately unbiased estimates. Notably, located latent class models (MMLLC) have slightly increased bias, but the efficiency with respect to RMSE is even higher than MMLN.

Table 3. Bias and relative RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 10$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution.

Method	Bias						Relative RMSE					
	NO	AM	UN	BE	LC2	LC3	NO	AM	UN	BE	LC2	LC3
MMLN	0.004	0.013	0.010	0.012	0.063	0.030	102.2	102.4	102.3	103.6	128.9	108.8
MMLLS(3)	0.003	0.006	0.007	0.010	0.029	0.015	101.6	101.0	101.5	102.7	112.2	103.9
MMLLS(4)	0.003	0.005	0.004	0.004	0.009	0.004	101.6	101.0	100.4	101.0	104.6	100.0
MMLMN(5)	0.008	0.005	0.005	0.005	0.023	0.011	101.0	100.4	100.6	102.4	106.2	105.3
MMLMN(7)	0.003	0.004	0.005	0.006	0.024	0.013	101.6	101.0	100.9	102.6	106.4	105.9
MMLMN(11)	0.003	0.004	0.005	0.005	0.006	0.005	101.7	101.2	100.2	100.5	100.0	101.1
MMLMN(15)	0.002	0.004	0.003	0.003	0.009	0.003	101.7	101.2	100.6	101.2	104.8	100.0
MMLLC(2)	0.054	0.055	0.044	0.037	0.019	0.033	118.5	118.0	111.1	108.1	102.8	106.0
MMLLC(3)	0.016	0.017	0.020	0.020	0.012	0.020	102.3	101.9	102.0	102.8	102.0	102.2
MMLLC(4)	0.009	0.011	0.012	0.011	0.007	0.008	101.7	101.2	101.0	101.4	101.8	100.5
MMLLC(5)	0.009	0.010	0.011	0.010	0.006	0.009	101.7	101.2	100.7	101.2	101.7	100.5
CML	0.004	0.003	0.003	0.004	0.003	0.004	102.6	101.8	101.2	101.9	102.6	101.4
JMLM	0.083	0.081	0.081	0.082	0.081	0.082	144.7	142.4	144.0	144.8	143.9	145.0
JMLW	0.036	0.034	0.037	0.039	0.036	0.038	113.4	111.8	113.5	114.8	113.5	114.3
PJML(1.0)	0.106	0.107	0.114	0.115	0.118	0.114	154.1	154.8	161.1	161.6	167.6	162.7
PJML(1.5)	0.002	0.010	0.009	0.011	0.047	0.024	100.0	100.0	100.0	101.0	116.6	103.9
PJML(2.0)	0.075	0.074	0.070	0.070	0.083	0.074	136.4	134.7	134.0	134.3	142.7	136.4
JML ϵ (0.1)	0.152	0.150	0.150	0.151	0.150	0.150	203.9	201.0	203.7	204.0	203.6	204.6
JML ϵ (0.2)	0.042	0.042	0.043	0.042	0.047	0.041	111.0	109.7	113.3	110.5	116.7	112.2
JML ϵ (0.24)	0.033	0.032	0.028	0.027	0.044	0.032	102.4	101.2	102.9	100.0	109.4	103.2
JML ϵ (0.3)	0.056	0.058	0.055	0.055	0.069	0.059	121.2	120.5	121.0	119.4	128.0	121.4
JML ϵ (0.4)	0.139	0.139	0.140	0.141	0.146	0.142	187.1	185.9	189.1	187.2	193.2	190.0
JML ϵ (0.5)	0.216	0.217	0.219	0.219	0.222	0.219	260.8	260.5	266.4	265.2	268.3	266.4
PMML	0.004	0.015	0.010	0.013	0.073	0.034	102.3	102.8	102.4	103.8	136.3	110.8
PCML	0.004	0.004	0.004	0.004	0.004	0.005	107.4	107.9	106.4	106.9	107.9	106.4
LLLA	0.004	0.014	0.010	0.013	0.066	0.031	102.0	102.2	102.2	103.6	130.4	109.2
MINCHI	0.003	0.004	0.004	0.003	0.003	0.002	106.5	107.1	105.7	106.1	107.1	105.4
EVM(2)	0.004	0.005	0.004	0.005	0.004	0.006	113.6	114.4	113.2	113.6	114.5	113.3
EVM(3)	0.004	0.005	0.004	0.005	0.004	0.006	114.0	114.9	113.7	113.9	114.9	113.7
RA(1)	0.019	0.020	0.018	0.020	0.019	0.021	126.9	128.5	127.2	127.2	127.6	128.0
RA(2)	0.004	0.005	0.004	0.005	0.004	0.006	113.6	114.4	113.2	113.6	114.5	113.3
RA(3)	0.004	0.005	0.004	0.005	0.004	0.006	114.0	114.9	113.7	113.9	114.9	113.7

Note. CML = conditional maximum likelihood; EVM = eigenvector method; JML = joint maximum likelihood; JML ϵ = JML with ϵ adjustment; JMLM = JML with maximum likelihood ability estimator; JMLW = JML with Warm's maximum likelihood ability estimator; LLLA = log-linear by linear association method; MINCHI = minimum chi-square estimation; MML = marginal maximum likelihood; MMLLC = MML with located latent classes; MMLLS = MML with log-linear smoothing; MMLMN = MML with multinomial distribution; MMLN = MML with normal distribution; PJML = penalized JML; PMML = pairwise MML; PCML = pairwise CML; RA = row-averaging method; NO = normal distribution; AM = asymmetric mixture distribution; UN = uniform distribution; BE = U-shaped beta distribution; LC2 = located 2-class distribution; LC3 = located 3-class distribution; Biases smaller than 0.025 and RMSE values smaller than 107 are printed in bold.

LIM were unbiased or had only small biases, except in the case in which θ was generated with located latent classes and PMML and LLLA estimation. This finding could be explained by the fact that these two estimation methods rely on the incorrect normal distribution assumption. Moreover, note that using powers 2 or 3 of the incidence matrix in RA (and EVM) improved estimates for bias and (to a larger extent) RMSE. Estimation methods PCML and MINCHI outperformed EVM and RA in terms of the RMSE. The additional computational burden in the iterative methods PCML and MINCHI compared to EVM and RA might be acceptable in practical applications.

The results in Table 3 also indicate that flexible MML estimation methods are competitive to CML estimation for nonnormally distributed abilities. Among the JML estimation

methods, the bias of PJML with prior $\sigma_{\text{prior}} = 1.5$ (i.e., PJML(1.5)) was smallest. Across data-generating models, the RMSE for this estimation method was smallest in three of the six data constellations, while in the other three constellations, JML ϵ with $\epsilon = 0.24$ performed best. However, it should also be emphasized that JML ϵ (0.24) introduced non-negligible bias in item parameters. The JMLW estimation method is preferred over JMLM in terms of bias and RMSE, but both methods were strongly inferior to PJML and JML ϵ .

Table A1 in Appendix B shows the bias and the RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 10$ items for a test with a small range of symmetrically distributed item difficulties as a function of the data-generating trait distribution. In general, biases in estimated item parameters were smaller than those with a wide range of item difficulties (presented in Table 3). This finding illustrates that item parameters with large true $|b_i|$ values (i.e., extremely small or extremely large) were more prone to bias than item difficulties with true parameters close to zero. It is also evident that the difference between the CML and MML methods from LIM was much smaller. Consequently, practitioners might prefer LIM for tests in which item difficulties do not have extreme values. In this case, the difference between PCML and MINCHI from RA and EVM also turned out to be smaller, although the former two methods might also be preferred. Interestingly and in contrast to a test with a wide range of item difficulties, JMLM outperformed JMLW. Moreover, note that PJML(1.5) resulted in less biased and more precise estimates than JML ϵ (0.24).

Table A2 in Appendix B shows the bias and the RMSE for different estimation methods for a sample size of $N = 250$ and $I = 10$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution. For a smaller sample size of $N = 250$ (compared to the large sample of $N = 1000$ in Table 3), JML ϵ (0.24) consistently resulted in the least RMSE but produced a slightly biased estimate. Surprisingly, more flexible MML estimation methods were not firmly inferior to MMLN if the θ followed a normal distribution. In particular, the located latent class model (MMLLC(2)) provided the most precise estimates among the MML methods for the sample size $N = 250$. Also, note that CML was not superior to all MML methods. Interestingly, the performance of LIM compared to MML and CML was even worse than for larger samples. Researchers should only probably opt for the computationally simpler LIM for sufficiently large sample sizes and tests with a smaller range of item difficulties.

Table 4 shows the bias and the RMSE for different estimation methods for a sample size of $N = 250$ and $I = 30$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution. For a longer test containing 30 items, biases for all estimation methods were substantially smaller than for 10 items. Notably, differences among estimation methods also turned out to be small. For example, assuming a misspecified normal distribution (i.e., MMLN) only introduced small biases. For a sufficiently long test, the differences between MML methods and CML from LIM were only modest, and practitioners might opt for the computationally simpler methods in this case. Again, PCML and MINCHI have slightly better performance than the EVM and RA methods. It is also important to note that JML ϵ required a larger ϵ value of 0.4 or 0.5 compared to a short test for realizing the maximum precision.

Table 4. Bias and relative RMSE for different estimation methods for a sample size of $N = 250$ and $I = 30$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution.

Method	Bias						Relative RMSE					
	NO	AM	UN	BE	LC2	LC3	NO	AM	UN	BE	LC2	LC3
MMLN	0.008	0.012	0.015	0.012	0.020	0.006	106.9	105.2	105.9	105.1	107.2	105.0
MMLLS(3)	0.007	0.008	0.014	0.011	0.011	0.009	106.7	104.4	105.7	105.1	105.8	105.4
MMLLS(4)	0.007	0.008	0.011	0.008	0.015	0.010	106.7	104.4	105.1	104.3	104.7	104.6
MMLMN(5)	0.011	0.012	0.009	0.009	0.030	0.030	106.2	104.2	104.5	103.4	105.5	107.0
MMLMN(7)	0.008	0.008	0.011	0.008	0.029	0.031	106.8	104.6	105.2	104.4	106.0	107.3
MMLMN(11)	0.008	0.008	0.011	0.007	0.027	0.014	106.7	104.5	105.1	104.3	105.5	105.6
MMLMN(15)	0.008	0.009	0.011	0.007	0.014	0.011	106.7	104.6	105.1	104.2	104.6	105.2
MMLLC(2)	0.029	0.025	0.017	0.014	0.004	0.018	103.8	102.0	102.6	102.1	104.2	102.4
MMLLC(3)	0.006	0.004	0.008	0.005	0.006	0.004	105.2	103.4	104.4	103.7	104.7	104.1
MMLLC(4)	0.005	0.006	0.010	0.007	0.007	0.005	106.1	104.1	104.9	104.0	104.8	104.6
MMLLC(5)	0.005	0.007	0.010	0.007	0.006	0.006	106.3	104.2	104.9	104.0	104.5	104.8
CML	0.008	0.009	0.012	0.008	0.009	0.008	106.9	104.8	105.3	104.4	105.6	105.1
JMLM	0.010	0.010	0.012	0.009	0.010	0.010	107.1	104.8	105.4	104.5	105.4	105.3
JMLW	0.007	0.005	0.007	0.007	0.004	0.007	105.2	103.1	103.5	102.7	104.0	103.3
PJML(1.0)	0.015	0.012	0.016	0.017	0.012	0.025	104.6	103.0	103.3	102.5	105.4	102.8
PJML(1.5)	0.010	0.013	0.014	0.010	0.021	0.005	107.1	105.4	105.6	104.7	107.6	104.7
PJML(2.0)	0.020	0.022	0.023	0.019	0.027	0.017	109.2	107.3	107.6	106.6	109.0	106.8
JML ϵ (0.1)	0.022	0.021	0.023	0.021	0.023	0.021	108.9	106.5	107.2	106.3	107.3	107.0
JML ϵ (0.2)	0.011	0.011	0.011	0.011	0.015	0.008	103.5	103.8	102.9	103.5	104.4	103.1
JML ϵ (0.24)	0.011	0.011	0.010	0.010	0.013	0.008	102.4	102.7	101.9	102.6	103.4	102.0
JML ϵ (0.3)	0.015	0.013	0.014	0.015	0.015	0.015	102.2	101.1	101.1	101.2	101.7	101.2
JML ϵ (0.4)	0.026	0.026	0.028	0.032	0.027	0.030	100.0	100.3	100.0	100.9	100.9	100.0
JML ϵ (0.5)	0.045	0.042	0.041	0.044	0.038	0.046	102.5	100.0	100.3	100.0	100.0	101.1
PMML	0.005	0.010	0.013	0.013	0.018	0.009	107.6	105.7	107.0	106.2	106.5	105.6
PCML	0.009	0.010	0.013	0.009	0.009	0.009	108.1	105.9	106.7	105.8	106.1	107.3
LLLA	0.013	0.016	0.019	0.016	0.019	0.014	107.6	105.8	106.6	105.7	106.9	106.1
MINCHI	0.005	0.006	0.009	0.006	0.006	0.005	107.2	104.9	105.7	104.8	105.3	106.4
EVM(2)	0.009	0.010	0.013	0.009	0.009	0.008	108.9	106.6	107.6	106.5	106.5	108.4
EVM(3)	0.009	0.010	0.013	0.009	0.009	0.009	108.9	106.6	107.6	106.5	106.4	108.4
RA(1)	0.018	0.019	0.021	0.017	0.017	0.019	111.2	108.8	109.9	108.9	108.4	110.9
RA(2)	0.009	0.010	0.013	0.009	0.009	0.008	108.9	106.6	107.6	106.5	106.5	108.4
RA(3)	0.009	0.010	0.013	0.009	0.009	0.009	108.9	106.6	107.6	106.5	106.4	108.4

Note. CML = conditional maximum likelihood; EVM = eigenvector method; JML = joint maximum likelihood; JML ϵ = JML with ϵ adjustment; JMLM = JML with maximum likelihood ability estimator; JMLW = JML with Warm's maximum likelihood ability estimator; LLLA = log-linear by linear association method; MINCHI = minimum chi-square estimation; MML = marginal maximum likelihood; MMLLLC = MML with located latent classes; MMLLS = MML with log-linear smoothing; MMLMN = MML with multinomial distribution; MMLN = MML with normal distribution; PJML = penalized JML; PMML = pairwise MML; PCML = pairwise CML; RA = row-averaging method; NO = normal distribution; AM = asymmetric mixture distribution; UN = uniform distribution; BE = U-shaped beta distribution; LC2 = located 2-class distribution; LC3 = located 3-class distribution; Biases smaller than 0.025 and RMSE values smaller than 107 are printed in bold.

Finally, Table A3 in Appendix B shows the bias and the RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 30$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution. Again, flexible MML specifications can compete with JML. PJML and JML ϵ (0.24) produced highly precise estimates, while the bias was almost negligible. It should be emphasized that these JML variants are at least as efficient as CML or MML variants. LIM resulted in more variable estimates. However, PCML and MINCHI almost achieve optimal efficiency of JML ϵ or MML variants. Surprisingly, PMML produced biased item parameter estimates and might not be recommended. Further research is needed whether this observation relied on the particular implementation of the authors.

5. Discussion

In this article, we compared several estimation methods for the Rasch model. It has been shown that the choice of the ability distribution impacts the precision of estimated item parameters. The differences between estimation methods appear larger for shorter (i.e., 10 items) than for longer (i.e., 30 items) tests. It turned out that MML with a flexible distribution can handle a nonnormally distributed trait well and can compete with CML. Interestingly, JML variants PJML and JML ϵ outperformed conditional and marginal maximum likelihood as well as LIM in many situations in terms of the RMSE. Moreover, these improved JML methods resulted in approximately unbiased estimates for long tests and larger sample sizes. These findings could stimulate research to consider JML methods PJML and JML ϵ instead of the widely implemented JMLM or JMLW variants. LIM are attractive for practitioners because they are not computationally demanding. It turned out that PCML or the MINCHI method outperformed the more widely used EVM or RA estimation methods.

Future research could investigate item parameter estimation in the RM for very short scales (e.g., $I = 5$ or $I = 7$ items). We suppose that differences among methods will appear larger in this situation. Moreover, optimal tuning parameter σ_{prior} in PJML and ϵ in JML ϵ depending on sample size, number of items, and item difficulty distribution have to be determined. We expect that optimal tuning parameters for individual ability estimates do not necessarily coincide with those that are optimal with respect to the RMSE of estimated parameters.

Throughout the simulation study, we assumed that the RM holds in the data. However, there might be situations where RM is intentionally a misspecified IRT model [129]. First, the two-parameter logistic model [130] might have generated the item responses, but the misspecified RM is used as a fitting model. In the case of misspecified IRT models, different estimation functions differently quantify model deviations (see also [110]). Future research might evaluate the appropriateness of estimation methods with respect to robustness from the assumption of the 1PL model. Notably, any estimation method defines its own set of item difficulties in the population of students because estimated difficulties are determined by a particular discrepancy function between the posed misspecified RM and a true IRT model that might involve very complex item response functions. Second, local dependence [34] is also often found in empirical data. LIM MINCHI, PCML, EVM and RA only rely on bivariate frequencies and not marginal frequencies. According to preliminary experience of the authors of this paper, these methods can result in less biased item parameter estimates than CML, MML methods, or JML methods. Studying the effects of local dependence for different estimation methods in the RM might be an exciting topic for future research.

In many applications, missing item responses often occur [131–133]. It can be expected that the studied estimation methods in this article can also be applied in situations in which data is missing completely at random (MCAR). It would be interesting whether MML has increased efficiency for MCAR data compared to LIM. For missing at random data, LIM (and ordinary CML) will likely fail (see [134]), and MML or JML will usually be preferred (see [87,135]).

Finally, we only considered frequentist estimation methods. In Bayesian estimation, prior distributions of item parameters can be included in the analysis, which can further stabilize the estimation of item difficulties [136–141]. We would like to note that when priors are normally distributed, a penalty function is added (or subtracted) in the estimation function. These penalties can be used not only for MML or JML but also CML [142] or LIM [143].

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CML	conditional maximum likelihood
EVM	eigenvector method
IRT	item response theory
JML	joint maximum likelihood
JML ϵ	JML with ϵ adjustment
JMLM	JML with maximum likelihood ability estimator
JMLW	JML with Warm's maximum likelihood ability estimator
LIM	limited information methods
LLLA	log-linear by linear association method
MAB	mean absolute bias
MCAR	missing completely at random
MML	marginal maximum likelihood
MMLLLC	MML with located latent classes
MMLLS	MML with log-linear smoothing
MMLMN	MML with multinomial distribution
MMLN	MML with normal distribution
MINCHI	minimum chi-square estimation
PJML	penalized JML
PMML	pairwise MML
PCML	pairwise CML
RA	row-averaging method
RM	Rasch model
RMSE	root mean square error

Appendix A. Item Parameters Used in the Simulation Study

The following sets of 10 item difficulties were used in the simulation study:

wide range of difficulties, symmetric difficulty distribution:

−3.000, −2.333, −1.667, −1.000, −0.333, 0.333, 1.000, 1.667, 2.333, 3.000

wide range of difficulties, asymmetric difficulty distribution:

−2.111, −2.037, −1.815, −1.444, −0.926, −0.259, 0.555, 1.518, 2.630, 3.889

small range of difficulties, symmetric difficulty distribution:

−1.500, −1.167, −0.833, −0.500, −0.167, 0.167, 0.500, 0.833, 1.167, 1.500

small range of difficulties, asymmetric difficulty distribution:

−1.055, −1.019, −0.907, −0.722, −0.463, −0.130, 0.278, 0.759, 1.315, 1.945

In the simulation condition with 30 items, each of the item difficulties is used three times. For example:

wide range of difficulties, symmetric difficulty distribution:

−3.000, −2.333, −1.667, −1.000, −0.333, 0.333, 1.000, 1.667, 2.333, 3.000, −3.000, −2.333, −1.667, −1.000, −0.333, 0.333, 1.000, 1.667, 2.333, 3.000, −3.000, −2.333, −1.667, −1.000, −0.333, 0.333, 1.000, 1.667, 2.333, 3.000

Appendix B. Additional Results for Simulation Study

Table A1 shows the bias and the RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 10$ items for a test with a small range of symmetrically distributed item difficulties as a function of the data-generating trait distribution. Table A2 shows the bias and the RMSE for different estimation methods for a sample size of $N = 250$ and $I = 10$ items for a test with a wide range of symmetrically distributed item difficulties as a

function of the data-generating trait distribution. Table A3 shows the bias and the RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 30$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution.

Table A1. Bias and relative RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 10$ items for a test with a small range of symmetrically distributed item difficulties as a function of the data-generating trait distribution.

Method	Bias						Relative RMSE					
	NO	AM	UN	BE	LC2	LC3	NO	AM	UN	BE	LC2	LC3
MMLN	0.002	0.005	0.007	0.010	0.013	0.004	100.7	101.5	102.0	102.8	105.8	101.1
MMLLS(3)	0.003	0.002	0.002	0.004	0.004	0.004	100.3	100.4	100.9	101.3	104.7	101.1
MMLLS(4)	0.003	0.003	0.004	0.004	0.004	0.005	100.3	100.4	100.3	100.4	105.7	100.0
MMLMN(5)	0.017	0.021	0.022	0.026	0.030	0.016	102.8	105.2	105.4	108.6	130.5	102.5
MMLMN(7)	0.003	0.004	0.004	0.004	0.028	0.004	100.3	100.5	100.1	100.1	113.1	100.1
MMLMN(11)	0.004	0.004	0.005	0.003	0.003	0.005	100.4	100.4	100.2	100.2	100.0	101.2
MMLMN(15)	0.004	0.003	0.004	0.004	0.003	0.004	100.4	100.5	100.4	100.6	106.0	100.8
MMLLC(2)	0.032	0.026	0.023	0.019	0.005	0.028	104.4	102.4	101.6	100.8	102.5	102.9
MMLLC(3)	0.015	0.013	0.013	0.012	0.004	0.016	100.3	100.1	100.1	100.1	102.6	100.3
MMLLC(4)	0.011	0.009	0.009	0.008	0.003	0.012	100.0	100.0	100.0	100.0	102.6	100.0
MMLLC(5)	0.010	0.008	0.008	0.007	0.004	0.011	100.0	100.1	100.1	100.0	102.4	100.1
CML	0.002	0.003	0.002	0.002	0.002	0.003	100.8	101.0	100.9	101.0	103.5	100.8
JMLM	0.015	0.014	0.014	0.013	0.015	0.014	104.3	103.9	104.0	103.7	106.2	104.2
JMLW	0.052	0.047	0.052	0.052	0.035	0.060	115.9	113.6	116.6	117.1	110.2	120.6
PJML(1.0)	0.045	0.035	0.045	0.044	0.009	0.063	112.1	107.7	112.7	113.0	103.7	123.1
PJML(1.5)	0.002	0.009	0.005	0.008	0.033	0.013	100.3	101.9	100.9	101.4	114.9	100.4
PJML(2.0)	0.030	0.037	0.032	0.032	0.053	0.021	112.0	114.9	112.1	112.6	129.9	107.0
JML ϵ (0.1)	0.047	0.049	0.048	0.048	0.052	0.044	122.2	122.6	122.1	122.4	128.7	120.1
JML ϵ (0.2)	0.027	0.027	0.024	0.022	0.030	0.025	104.4	102.9	102.3	102.8	107.8	100.6
JML ϵ (0.24)	0.036	0.035	0.032	0.032	0.034	0.038	107.4	105.4	104.6	105.0	108.4	104.6
JML ϵ (0.3)	0.060	0.058	0.057	0.056	0.053	0.063	120.5	117.8	117.6	117.5	117.5	120.0
JML ϵ (0.4)	0.101	0.099	0.097	0.096	0.088	0.105	153.4	150.0	149.2	149.9	146.3	153.7
JML ϵ (0.5)	0.139	0.135	0.137	0.135	0.124	0.144	189.9	184.7	186.5	186.5	179.7	191.9
PMML	0.002	0.005	0.007	0.011	0.012	0.004	100.8	101.3	101.9	102.7	105.3	101.3
PCML	0.002	0.003	0.003	0.003	0.002	0.003	103.2	102.7	103.0	102.6	105.1	102.6
LLLA	0.004	0.007	0.010	0.014	0.014	0.008	101.1	102.0	102.9	103.9	105.8	102.2
MINCHI	0.002	0.002	0.002	0.002	0.002	0.003	103.0	102.5	102.8	102.4	104.9	102.4
EVM(2)	0.002	0.003	0.002	0.003	0.002	0.003	104.4	103.5	104.1	103.4	106.0	103.6
EVM(3)	0.002	0.003	0.002	0.003	0.002	0.003	104.3	103.5	104.1	103.4	105.9	103.6
RA(1)	0.003	0.004	0.004	0.004	0.004	0.004	104.8	104.1	104.7	104.1	106.4	104.3
RA(2)	0.002	0.003	0.002	0.003	0.002	0.003	104.4	103.5	104.1	103.4	106.0	103.6
RA(3)	0.002	0.003	0.002	0.003	0.002	0.003	104.3	103.5	104.1	103.4	105.9	103.6

Note. CML = conditional maximum likelihood; EVM = eigenvector method; JML = joint maximum likelihood; JML ϵ = JML with ϵ adjustment; JMLM = JML with maximum likelihood ability estimator; JMLW = JML with Warm's maximum likelihood ability estimator; LLLA = log-linear by linear association method; MINCHI = minimum chi-square estimation; MML = marginal maximum likelihood; MMLLC = MML with located latent classes; MMLLS = MML with log-linear smoothing; MMLMN = MML with multinomial distribution; MMLN = MML with normal distribution; PJML = penalized JML; PMML = pairwise MML; PCML = pairwise CML; RA = row-averaging method; NO = normal distribution; AM = asymmetric mixture distribution; UN = uniform distribution; BE = U-shaped beta distribution; LC2 = located 2-class distribution; LC3 = located 3-class distribution; Biases smaller than 0.025 and RMSE values smaller than 107 are printed in bold.

Table A2. Bias and relative RMSE for different estimation methods for a sample size of $N = 250$ and $I = 10$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution.

Method	Bias						Relative RMSE					
	NO	AM	UN	BE	LC2	LC3	NO	AM	UN	BE	LC2	LC3
MMLN	0.014	0.021	0.020	0.018	0.070	0.032	108.6	104.0	107.0	108.0	116.1	109.0
MMLLS(3)	0.012	0.013	0.020	0.017	0.045	0.023	108.6	103.6	107.3	108.1	115.9	109.1
MMLLS(4)	0.011	0.011	0.017	0.014	0.024	0.015	108.5	103.4	106.3	106.7	110.1	106.8
MMLMN(5)	0.011	0.011	0.020	0.020	0.037	0.019	107.8	102.7	106.8	108.5	108.3	108.9
MMLMN(7)	0.015	0.014	0.022	0.021	0.038	0.022	109.0	104.0	107.6	109.0	109.1	110.7
MMLMN(11)	0.014	0.014	0.020	0.017	0.023	0.019	109.0	104.0	106.7	107.2	107.4	108.3
MMLMN(15)	0.015	0.015	0.019	0.016	0.026	0.017	109.0	104.0	106.8	107.1	110.7	107.2
MMLLC(2)	0.029	0.028	0.014	0.011	0.011	0.010	106.2	101.1	103.1	104.5	106.9	104.4
MMLLC(3)	0.007	0.008	0.013	0.009	0.020	0.010	107.9	103.0	105.8	106.3	108.5	106.1
MMLLC(4)	0.010	0.012	0.016	0.012	0.023	0.012	108.5	103.5	106.3	106.6	109.0	106.7
MMLLC(5)	0.011	0.013	0.016	0.012	0.022	0.013	108.7	103.7	106.4	106.8	108.9	106.8
CML	0.014	0.015	0.017	0.013	0.016	0.013	109.0	104.1	106.5	106.8	108.6	106.9
JMLM	0.093	0.095	0.097	0.094	0.096	0.091	128.8	123.7	127.3	126.7	128.5	126.8
JMLW	0.046	0.048	0.052	0.049	0.050	0.047	115.4	110.2	114.0	114.3	114.9	113.8
PJML(1.0)	0.099	0.098	0.103	0.107	0.110	0.107	116.9	112.5	115.4	117.9	121.1	118.4
PJML(1.5)	0.009	0.016	0.014	0.013	0.054	0.025	106.4	101.8	104.0	105.0	110.5	105.6
PJML(2.0)	0.084	0.085	0.083	0.079	0.095	0.082	122.5	117.7	120.2	120.0	124.0	120.4
JML ϵ (0.1)	0.160	0.162	0.163	0.160	0.162	0.158	150.0	145.1	148.8	148.0	149.0	148.1
JML ϵ (0.2)	0.053	0.051	0.050	0.050	0.047	0.047	107.0	106.5	106.5	106.9	105.1	106.4
JML ϵ (0.24)	0.034	0.035	0.032	0.027	0.043	0.029	100.0	100.0	100.0	100.0	100.0	100.0
JML ϵ (0.3)	0.049	0.053	0.049	0.050	0.066	0.056	102.7	100.8	101.3	101.8	104.0	102.5
JML ϵ (0.4)	0.132	0.133	0.135	0.136	0.146	0.140	123.4	122.5	123.8	123.9	126.0	125.2
JML ϵ (0.5)	0.213	0.211	0.213	0.216	0.218	0.217	156.9	152.5	154.1	156.6	155.8	157.6
PMML	0.014	0.022	0.020	0.019	0.081	0.035	108.7	104.2	107.1	108.1	118.9	109.7
PCML	0.017	0.018	0.021	0.016	0.019	0.016	115.2	109.8	112.8	113.2	114.4	112.3
LLLA	0.013	0.021	0.020	0.018	0.073	0.032	108.4	103.9	106.9	107.9	116.5	109.0
MINCHI	0.008	0.008	0.006	0.010	0.008	0.010	111.4	106.2	108.8	109.6	110.6	108.8
EVM(2)	0.020	0.021	0.024	0.019	0.022	0.018	122.9	117.2	120.9	121.4	122.6	119.9
EVM(3)	0.020	0.021	0.024	0.019	0.022	0.018	123.0	117.3	121.1	121.5	122.7	120.0
RA(1)	0.026	0.028	0.027	0.026	0.027	0.026	115.8	111.3	114.0	114.8	114.6	114.2
RA(2)	0.020	0.021	0.024	0.018	0.022	0.018	122.9	117.2	120.9	121.4	122.6	119.9
RA(3)	0.020	0.021	0.024	0.019	0.022	0.018	123.0	117.3	121.1	121.5	122.7	120.0

Note. CML = conditional maximum likelihood; EVM = eigenvector method; JML = joint maximum likelihood; JML ϵ = JML with ϵ adjustment; JMLM = JML with maximum likelihood ability estimator; JMLW = JML with Warm's maximum likelihood ability estimator; LLLA = log-linear by linear association method; MINCHI = minimum chi-square estimation; MML = marginal maximum likelihood; MMLLC = MML with located latent classes; MMLLS = MML with log-linear smoothing; MMLMN = MML with multinomial distribution; MMLN = MML with normal distribution; PJML = penalized JML; PMML = pairwise MML; PCML = pairwise CML; RA = row-averaging method; NO = normal distribution; AM = asymmetric mixture distribution; UN = uniform distribution; BE = U-shaped beta distribution; LC2 = located 2-class distribution; LC3 = located 3-class distribution; Biases smaller than 0.025 and RMSE values smaller than 107 are printed in bold.

Table A3. Bias and relative RMSE for different estimation methods for a sample size of $N = 1000$ and $I = 30$ items for a test with a wide range of symmetrically distributed item difficulties as a function of the data-generating trait distribution.

Method	Bias						Relative RMSE					
	NO	AM	UN	BE	LC2	LC3	NO	AM	UN	BE	LC2	LC3
MMLN	0.003	0.009	0.008	0.010	0.040	0.018	101.2	101.4	102.3	105.1	112.3	103.8
MMLLS(3)	0.003	0.003	0.007	0.009	0.035	0.011	100.4	100.2	101.6	104.7	110.6	101.9
MMLLS(4)	0.003	0.005	0.004	0.004	0.008	0.007	100.3	100.2	100.5	102.5	101.5	100.9
MMLMN(5)	0.010	0.006	0.006	0.005	0.024	0.016	100.4	100.0	100.5	103.6	107.6	102.2
MMLMN(7)	0.003	0.004	0.006	0.005	0.024	0.016	100.3	100.2	100.5	103.6	107.5	102.8
MMLMN(11)	0.003	0.004	0.005	0.005	0.009	0.013	100.3	100.2	100.4	102.4	100.9	101.3
MMLMN(15)	0.002	0.004	0.004	0.003	0.008	0.007	100.4	100.2	100.5	102.7	101.7	100.0
MMLLC(2)	0.083	0.082	0.065	0.052	0.016	0.040	136.7	135.7	123.3	117.0	101.4	108.5
MMLLC(3)	0.031	0.030	0.024	0.021	0.008	0.019	105.5	104.9	103.1	104.3	100.6	101.8
MMLLC(4)	0.013	0.012	0.014	0.013	0.006	0.009	100.9	100.7	101.2	103.2	100.6	100.0
MMLLC(5)	0.008	0.007	0.009	0.009	0.007	0.010	100.6	100.3	100.8	102.9	100.5	100.2
CML	0.003	0.003	0.005	0.005	0.004	0.002	101.3	101.0	101.4	103.7	101.3	100.6
JMLM	0.024	0.025	0.025	0.026	0.025	0.023	105.9	105.8	106.5	109.1	106.1	105.0
JMLW	0.015	0.015	0.016	0.018	0.016	0.014	103.1	102.9	103.7	106.3	103.1	102.5
PJML(1.0)	0.052	0.051	0.056	0.055	0.062	0.058	115.1	114.8	117.4	118.8	122.8	118.8
PJML(1.5)	0.007	0.008	0.008	0.008	0.024	0.011	101.4	101.3	102.0	104.6	105.2	101.9
PJML(2.0)	0.035	0.036	0.034	0.036	0.039	0.033	110.6	110.5	111.2	114.0	111.7	109.7
JML ϵ (0.1)	0.050	0.051	0.051	0.052	0.050	0.049	117.8	117.8	118.9	121.5	117.7	116.5
JML ϵ (0.2)	0.019	0.019	0.018	0.020	0.020	0.020	102.7	104.2	102.5	103.0	102.3	103.3
JML ϵ (0.24)	0.013	0.012	0.013	0.011	0.015	0.014	100.0	101.4	100.0	100.0	100.0	100.5
JML ϵ (0.3)	0.016	0.016	0.015	0.013	0.021	0.016	100.1	100.8	100.1	100.7	100.7	100.0
JML ϵ (0.4)	0.039	0.040	0.041	0.039	0.044	0.041	109.3	109.9	109.9	108.3	110.1	108.9
JML ϵ (0.5)	0.067	0.067	0.068	0.067	0.072	0.070	125.6	125.5	125.6	126.2	126.6	126.0
PMML	0.049	0.045	0.105	0.112	0.079	0.097	161.5	158.1	210.7	217.0	182.8	202.4
PCML	0.004	0.004	0.005	0.006	0.004	0.003	104.6	104.5	104.5	106.7	104.2	103.9
LLLA	0.003	0.010	0.008	0.010	0.042	0.019	101.0	101.3	102.2	105.0	113.2	103.9
MINCHI	0.004	0.004	0.004	0.002	0.003	0.005	103.9	103.8	103.7	105.8	103.4	103.4
EVM(2)	0.004	0.004	0.005	0.006	0.004	0.003	109.2	109.2	109.1	111.3	108.4	108.6
EVM(3)	0.004	0.004	0.005	0.006	0.004	0.003	109.2	109.2	109.2	111.3	108.5	108.7
RA(1)	0.020	0.020	0.020	0.022	0.019	0.019	117.1	117.0	117.2	120.0	115.5	116.5
RA(2)	0.004	0.004	0.005	0.006	0.004	0.003	109.2	109.2	109.1	111.3	108.4	108.6
RA(3)	0.004	0.004	0.005	0.006	0.004	0.003	109.2	109.2	109.2	111.3	108.5	108.7

Note. CML = conditional maximum likelihood; EVM = eigenvector method; JML = joint maximum likelihood; JML ϵ = JML with ϵ adjustment; JMLM = JML with maximum likelihood ability estimator; JMLW = JML with Warm's maximum likelihood ability estimator; LLLA = log-linear by linear association method; MINCHI = minimum chi-square estimation; MML = marginal maximum likelihood; MMLLLC = MML with located latent classes; MMLLS = MML with log-linear smoothing; MMLMN = MML with multinomial distribution; MMLN = MML with normal distribution; PJML = penalized JML; PMML = pairwise MML; PCML = pairwise CML; RA = row-averaging method; NO = normal distribution; AM = asymmetric mixture distribution; UN = uniform distribution; BE = U-shaped beta distribution; LC2 = located 2-class distribution; LC3 = located 3-class distribution; Biases smaller than 0.025 and RMSE values smaller than 107 are printed in bold.

References

1. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Danish Institute for Educational Research: Copenhagen, Denmark, 1960.
2. Fischer, G.H.; Molenaar, I.W. *Rasch Models. Foundations, Recent Developments, and Applications*; Springer: New York, NY, USA, 1995. doi:10.1007/978-1-4612-4230-7.
3. von Davier, M. The Rasch model. In *Handbook of Item Response Theory, Volume 1: Models*; CRC Press: Boca Raton, FL, USA, 2016; pp. 31–48. doi:10.1201/9781315119144.
4. Baker, F.B.; Kim, S.H. *Item Response Theory: Parameter Estimation Techniques*; CRC Press: Boca Raton, FL, USA, 2004. doi:10.1201/9781482276725.
5. Cai, L.; Choi, K.; Hansen, M.; Harrell, L. Item response theory. *Annu. Rev. Stat. Appl.* **2016**, *3*, 297–321, doi:10.1146/annurev-statistics-041715-033702.

6. Chen, Y.; Li, X.; Liu, J.; Ying, Z. Item response theory—A statistical framework for educational and psychological measurement. *arXiv* **2021**, arXiv:2108.08604. <https://arxiv.org/abs/2108.08604>.
7. van der Linden, W.J.; Hambleton, R.K. *Handbook of Modern Item Response Theory*; Springer: New York, NY, USA, 1997. doi:10.1007/978-1-4757-2691-6.
8. Lord, F.M.; Novick, R. *Statistical Theories of Mental Test Scores*; Addison-Wesley: Reading, MA, USA, 1968.
9. Yen, W.M.; Fitzpatrick, A.R. Item response theory. In *Educational Measurement*; Brennan, R.L., Ed.; Praeger: New York, NY, USA, 2006; pp. 111–154.
10. Arnold, J.C.; Boone, W.J.; Kremer, K.; Mayer, J. Assessment of competencies in scientific inquiry through the application of Rasch measurement techniques. *Educ. Sci.* **2018**, *8*, 184. doi:10.3390/educsci8040184.
11. Cascella, C.; Giberti, C.; Bolondi, G. Changing the order of factors does not change the product but does affect students' answers, especially girls' answers. *Educ. Sci.* **2021**, *11*, 201. doi:10.3390/educsci11050201.
12. Finger, M.E.; Escorpizo, R.; Tennant, A. Measuring work-related functioning using the work rehabilitation questionnaire (WORQ). *Int. J. Environ. Res. Public Health* **2019**, *16*, 2795. doi:10.3390/ijerph16152795.
13. Kramer, M.; Förtsch, C.; Boone, W.J.; Seidel, T.; Neuhaus, B.J. Investigating pre-service biology teachers' diagnostic competences: Relationships between professional knowledge, diagnostic activities, and diagnostic accuracy. *Educ. Sci.* **2021**, *11*, 89. doi:10.3390/educsci11030089.
14. Morales-Rodríguez, F.M.; Martí-Vilar, M.; Peláez, M.A.N.; Lozano, J.M.G.; Ramón, J.P.M.; Caracuel, A. Psychometric properties of the affective dimension of the generic macro-competence assessment scale: Analysis using Rasch model. *Sustainability* **2021**, *13*, 6904. doi:10.3390/su13126904.
15. Raccanello, D.; Vicentini, G.; Burro, R. Children's psychological representation of earthquakes: Analysis of written definitions and Rasch scaling. *Geosciences* **2019**, *9*, 208. doi:10.3390/geosciences9050208.
16. Shoahosseini, R.; Baghaei, P. Validation of the Persian translation of the children's test anxiety scale: A multidimensional Rasch model analysis. *Eur. J. Investig. Health Psychol. Educ.* **2020**, *10*, 59–69. doi:10.3390/ejihpe10010006.
17. Andrich, D.; Marais, I. *A Course in Rasch Measurement Theory*; Springer: New York, NY, USA, 2019. doi:10.1007/978-981-13-7496-8.
18. Boone, W.J. Rasch analysis for instrument development: Why, when, and how? *CBE Life Sci. Educ.* **2016**, *15*, rm4. doi:10.1187/cbe.16-04-0148.
19. Bond, T.; Yan, Z.; Heene, M. *Applying the Rasch Model*; Routledge: New York, NY, USA, 2020. doi:10.4324/9780429030499.
20. Engelhard, G. *Invariant Measurement*; Routledge: New York, NY, USA, 2012. doi:10.4324/9780203073636.
21. Lamprianou, I. *Applying the Rasch Model in Social Sciences Using R and BlueSky Statistics*; Routledge: New York, NY, USA, 2019. doi:10.4324/9781315146850.
22. Linacre, J.M. Understanding Rasch measurement: estimation methods for Rasch measures. *J. Outcome Meas.* **1999**, *3*, 382–405.
23. Linacre, J.M. Rasch model estimation: Further topics. *J. Appl. Meas.* **2004**, *5*, 95–110.
24. Wilson, M. *Constructing Measures: An Item Response Modeling Approach*; Routledge: New York, NY, USA, 2004. doi:10.4324/9781410611697.
25. Wright, B.D.; Stone, M.H. *Best Test Design*; Mesa Press: Chicago, IL, USA, 1979.
26. Wu, M.; Tam, H.P.; Jen, T.H. *Educational Measurement for Applied Researchers*; Springer: Singapore, 2016. doi:10.1007/978-981-10-3302-5.
27. Aryadoust, V.; Tan, H.A.H.; Ng, L.Y. A Scientometric review of Rasch measurement: The rise and progress of a specialty. *Front. Psychol.* **2019**, *10*, 2197. doi:10.3389/fpsyg.2019.02197.
28. De Boeck, P. Random item IRT models. *Psychometrika* **2008**, *73*, 533–559. doi:10.1007/s11336-008-9092-x.
29. Holland, P.W. On the sampling theory foundations of item response theory models. *Psychometrika* **1990**, *55*, 577–601. doi:10.1007/BF02294609.
30. Fischer, G.H. Rasch models. In *Handbook of Statistics, Vol. 26: Psychometrics*; Rao, C.R., Sinharay, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 515–585. doi:10.1016/S0169-7161(06)26016-4.
31. Wu, M.; Adams, R.J. Properties of Rasch residual fit statistics. *J. Appl. Meas.* **2013**, *14*, 339–355.
32. Christensen, K.B.; Makransky, G.; Horton, M. Critical values for Yen's Q_3 : Identification of local dependence in the Rasch model using residual correlations. *Appl. Psychol. Meas.* **2017**, *41*, 178–194. doi:10.1177/0146621616677520.
33. Debelak, R.; Koller, I. Testing the local independence assumption of the Rasch model with Q_3 -based nonparametric model tests. *Appl. Psychol. Meas.* **2020**, *44*, 103–117. doi:10.1177/0146621619835501.
34. Yen, W.M. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Appl. Psychol. Meas.* **1984**, *8*, 125–145. doi:10.1177/014662168400800201.
35. Meyer, P. *Understanding Measurement: Reliability*; Oxford University Press: Cambridge, UK, 2010. doi:10.1093/acprof:oso/9780195380361.001.0001.
36. Fan, X. Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educ. Psychol. Meas.* **1998**, *58*, 357–381. doi:10.1177/0013164498058003001.
37. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. doi:10.18637/jss.v067.i01.
38. De Boeck, P.; Wilson, M. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*; Springer: New York, NY, USA, 2004. doi:10.1007/978-1-4757-3990-9.

39. Doran, H.; Bates, D.; Bliese, P.; Dowling, M. Estimating the multilevel Rasch model: With the lme4 package. *J. Stat. Softw.* **2007**, *20*, 1–18. doi:10.18637/jss.v020.i02.
40. Rijmen, F.; Tuerlinckx, F.; De Boeck, P.; Kuppens, P. A nonlinear mixed model framework for item response theory. *Psychol. Methods* **2003**, *8*, 185–205. doi:10.1037/1082-989X.8.2.185.
41. Zheng, X.; Rabe-Hesketh, S. Estimating parameters of dichotomous and ordinal item response models with gglamm. *Stata J.* **2007**, *7*, 313–333. doi:10.1177/1536867X0700700302.
42. Raudenbush, S.W.; Johnson, C.; Sampson, R.J. A multivariate, multilevel Rasch model with application to self-reported criminal behavior. *Sociol. Methodol.* **2003**, *33*, 169–211. doi:10.1111/j.0081-1750.2003.t01-1-00130.x.
43. Molenaar, I.W. Estimation of item parameters. In *Rasch Models. Foundations, Recent Developments, and Applications*; Fischer, G.H., Molenaar, I.W., Eds.; Springer: New York, NY, USA, 1995; pp. 39–52. doi:10.1007/978-1-4612-4230-7_3.
44. Wainer, H.; Morgan, A.; Gustafsson, J.E. A review of estimation procedures for the Rasch model with an eye toward longish tests. *J. Educ. Stat.* **1980**, *5*, 35–64. doi:10.3102/10769986005001035.
45. Haberman, S.J. *Joint and Conditional Maximum Likelihood Estimation for the Rasch Model for Binary Responses*; (Research Report No. RR-04-20); Educational Testing Service: Princeton, NJ, USA, 2004. doi:10.1002/j.2333-8504.2004.tb01947.x.
46. Haberman, S.J. Maximum likelihood estimates in exponential response models. *Ann. Stat.* **1977**, *5*, 815–841. doi:10.1214/aos/1176343941.
47. Haberman, S.J. Models with nuisance and incidental parameters. In *Handbook of Item Response Theory, Vol. 2: Statistical Tools*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 151–170. doi:10.1201/b19166-9.
48. Lancaster, T. The incidental parameter problem since 1948. *J. Econom.* **2000**, *95*, 391–413. doi:10.1016/S0304-4076(99)00044-5.
49. Warm, T.A. Weighted likelihood estimation of ability in item response theory. *Psychometrika* **1989**, *54*, 427–450. doi:10.1007/BF02294627.
50. Magis, D.; Raiche, G. On the relationships between Jeffreys modal and weighted likelihood estimation of ability under logistic IRT models. *Psychometrika* **2012**, *77*, 163–169. doi:10.1007/s11336-011-9233-5.
51. Jansen, P.G.W.; van den Wollenberg, A.L.; Wierda, F.W. Correcting unconditional parameter estimates in the Rasch model for inconsistency. *Appl. Psychol. Meas.* **1988**, *12*, 297–306. doi:10.1177/014662168801200307.
52. Wright, B.D.; Douglas, G.A. Best procedures for sample-free item analysis. *Appl. Psychol. Meas.* **1977**, *1*, 281–295. doi:10.1177/014662167700100216.
53. Chen, Y.; Li, X.; Zhang, S. Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika* **2019**, *84*, 124–146. doi:10.1007/s11336-018-9646-5.
54. Paolino, J.P. *Penalized Joint Maximum Likelihood Estimation Applied to Two Parameter Logistic Item Response Models*. Ph.D. Thesis, Columbia University, New York, NY, USA, 2013. doi:10.7916/D88W3MHS.
55. Paolino, J.P. Rasch model parameter estimation via the elastic net. *J. Appl. Meas.* **2015**, *16*, 353–364.
56. Bertoli-Barsotti, L.; Lando, T.; Punzo, A. Estimating a Rasch Model via Fuzzy Empirical Probability Functions. In *Analysis and Modeling of Complex Data in Behavioral and Social Sciences*; Vicari, D., Okada, A., Ragozini, G., Weihs, C., Eds.; Springer: Cham, Switzerland, 2014; pp. 29–36. doi:10.1007/978-3-319-06692-9_4.
57. Lando, T.; Bertoli-Barsotti, L. A modified minimum divergence estimator: some preliminary results for the Rasch model. *Electr. J. Appl. Stat. Anal.* **2014**, *7*, 37–57. doi:10.1285/i20705948v7n1p37.
58. Robitzsch, A.; Steinfeld, J. Item response models for human ratings: Overview, estimation methods, and implementation in R. *Psych. Test Assess. Model.* **2018**, *60*, 101–138.
59. Andersen, E.B. The numerical solution of a set of conditional estimation equations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1972**, *34*, 42–54. doi:10.1111/j.2517-6161.1972.tb00887.x.
60. Draxler, C.; Alexandrowicz, R.W. Sample size determination within the scope of conditional maximum likelihood estimation with special focus on testing the Rasch model. *Psychometrika* **2015**, *80*, 897–919. doi:10.1007/s11336-015-9472-y.
61. Mair, P.; Hatzinger, R. CML based estimation of extended Rasch models with the eRm package in R. *Psychol. Sci.* **2007**, *49*, 26–43.
62. Hatzinger, R.; Rusch, T. IRT models with relaxed assumptions in eRm: A manual-like instruction. *Psychol. Sci. Q.* **2009**, *51*, 87–120.
63. Liou, M. More on the computation of higher-order derivatives of the elementary symmetric functions in the Rasch model. *Appl. Psychol. Meas.* **1994**, *18*, 53–62. doi:10.1177/014662169401800105.
64. Verhelst, N.D.; Glas, C.A.W.; Van der Sluis, A. Estimation problems in the Rasch model: The basic symmetric functions. *Comp. Stat. Q.* **1984**, *1*, 245–262.
65. Bartolucci, F.; Pigini, C. cquad: An R and Stata package for conditional maximum likelihood estimation of dynamic binary panel data models. *J. Stat. Softw.* **2017**, *78*, 1–26. doi:10.18637/jss.v078.i07.
66. Duchesne, T.; Fortin, D.; Courbin, N. Mixed conditional logistic regression for habitat selection studies. *J. Anim. Ecol.* **2010**, *79*, 548–555. doi:10.1111/j.1365-2656.2010.01670.x.
67. Sartori, N.; Severini, T.A. Conditional likelihood inference in generalized linear mixed models. *Stat. Sin.* **2004**, *14*, 349–360.
68. Aitkin, M. Expectation maximization algorithm and extensions. In *Handbook of Item Response Theory, Vol. 2: Statistical Tools*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 217–236. doi:10.1201/b19166-12.
69. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459. doi:10.1007/BF02293801.

70. Abdel-Fattah, A. Comparing BILOG and LOGIST estimates for normal, truncated normal, and beta ability distributions. In Proceedings of the Annual Meeting of the American Educational Research Association, New Orleans, LA, USA, 4–8 April 1994.
71. Woods, C.M. Estimating the latent density in unidimensional IRT to permit non-normality. In *Handbook of Item Response Theory Modeling*; Reise, S.P., Revicki, D.A., Eds.; Routledge: New York, NY, USA, 2014; pp. 78–102. doi:10.4324/9781315736013.
72. Chalmers, R.P. mirt: A multidimensional item response theory package for the R environment. *J. Stat. Softw.* **2012**, *48*, 1–29. doi:10.18637/jss.v048.i06.
73. Robitzsch, A.; Kiefer, T.; Wu, M. TAM: Test Analysis Modules; R Package Version 3.7-6. 2021. Available online: <https://CRAN.R-project.org/package=TAM> (accessed on 25 June 2021).
74. Kirisci, L.; Hsu, T.C.; Yu, L. Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Appl. Psychol. Meas.* **2001**, *25*, 146–162. doi:10.1177/01466210122031975.
75. Seong, T.J. Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Appl. Psychol. Meas.* **1990**, *14*, 299–311. doi:10.1177/014662169001400307.
76. Stone, C.A. Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Appl. Psychol. Meas.* **1992**, *16*, 1–16. doi:10.1177/014662169201600101.
77. Zwinderman, A.H.; Van den Wollenberg, A.L. Robustness of marginal maximum likelihood estimation in the Rasch model. *Appl. Psychol. Meas.* **1990**, *14*, 73–81. doi:10.1177/014662169001400107.
78. Grilli, L.; Metelli, S.; Rampichini, C. Bayesian estimation with integrated nested Laplace approximation for binary logit mixed models. *J. Stat. Comput. Simul.* **2015**, *85*, 2718–2726. doi:10.1080/00949655.2014.935377.
79. Hedeker, D. A mixed-effects multinomial logistic regression model. *Stat. Med.* **2003**, *22*, 1433–1446. doi:10.1002/sim.1522.
80. Raudenbush, S.W.; Yang, M.L.; Yosef, M. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *J. Comput. Graph. Stat.* **2000**, *9*, 141–157. doi:10.2307/1390617.
81. Woods, C.M. Empirical histograms in item response theory with ordinal data. *Educ. Psychol. Meas.* **2007**, *67*, 73–87. doi:10.1177/0013164406288163.
82. von Davier, M. A general diagnostic model applied to language testing data. *Br. J. Math. Stat. Psychol.* **2008**, *61*, 287–307. doi:10.1348/000711007X193957.
83. Xu, X.; von Davier, M. *Fitting the Structured General Diagnostic Model to NAEP Data*; (Research Report No. RR-08-28); Educational Testing Service: Princeton, NJ, USA, 2008. doi:10.1002/j.2333-8504.2008.tb02113.x.
84. Casabianca, J.M.; Junker, B.W. Estimating the latent trait distribution with loglinear smoothing models. In *New Developments in Quantitative Psychology*; Millsap, R.E., van der Ark, L.A., Bolt, D.M., Woods, C.M., Eds.; Springer: New York, NY, USA, 2013; pp. 415–425. doi:10.1007/978-1-4614-9348-8_27.
85. Casabianca, J.M.; Lewis, C. IRT item parameter recovery with marginal maximum likelihood estimation using loglinear smoothing models. *J. Educ. Behav. Stat.* **2015**, *40*, 547–578. doi:10.3102/1076998615606112.
86. Haberman, S.J.; von Davier, M.; Lee, Y.H. *Comparison of Multidimensional Item Response Models: Multivariate Normal Ability Distributions Versus Multivariate Polytomous Distributions*; (Research Report No. RR-08-45); Educational Testing Service: Princeton, NJ, USA, 2008. doi:10.1002/j.2333-8504.2008.tb02131.x.
87. Steinfeld, J.; Robitzsch, A. Item parameter estimation in multistage designs: A comparison of different estimation approaches for the Rasch model. *Psych* **2021**, *3*, 279–307. doi:10.3390/psych3030022.
88. Xu, X.; von Davier, M. *Comparing Multiple-Group Multinomial Log-Linear Models for Multidimensional Skill Distributions in the General Diagnostic Model*; (Research Report No. RR-08-35); Educational Testing Service: Princeton, NJ, USA, 2008. doi:10.1002/j.2333-8504.2008.tb02121.x.
89. De Leeuw, J.; Verhelst, N. Maximum likelihood estimation in generalized Rasch models. *J. Educ. Behav. Stat.* **1986**, *11*, 183–196. doi:10.3102/10769986011003183.
90. Formann, A.K. Constrained latent class models: Theory and applications. *Br. J. Math. Stat. Psychol.* **1985**, *38*, 87–111. doi:10.1111/j.2044-8317.1985.tb00818.x.
91. Haberman, S.J. *Latent-Class Item Response Models*; (Research Report No. RR-05-28); Educational Testing Service: Princeton, NJ, USA, 2005. doi:10.1002/j.2333-8504.2005.tb02005.x.
92. Lindsay, B.; Clogg, C.C.; Grego, J. Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Am. Stat. Assoc.* **1991**, *86*, 96–107. doi:10.2307/2289719.
93. Bacci, S.; Bartolucci, F. A multidimensional latent class Rasch model for the assessment of the health-related quality of life. In *Rasch Models in Health*; Christensen, K.B., Kreiner, S., Mesbah, M., Eds.; Wiley: Hoboken, NJ, USA, 2013; pp. 197–218. doi:10.1002/9781118574454.ch11.
94. Genge, E. LC and LC-IRT models in the identification of Polish households with similar perception of financial position. *Sustainability* **2021**, *13*, 4130. doi:10.3390/su13084130.
95. Katsikatsou, M.; Moustaki, I.; Yang-Wallentin, F.; Jöreskog, K.G. Pairwise likelihood estimation for factor analysis models with ordinal data. *Comput. Stat. Data Anal.* **2012**, *56*, 4243–4258. doi:10.1016/j.csda.2012.04.010.
96. Feddag, M.L.; Hardouin, J.B.; Seville, V. Pairwise- and marginal-likelihood estimation for the mixed Rasch model with binary data. *J. Stat. Comput. Simul.* **2012**, *82*, 419–430. doi:10.1080/00949655.2010.538691.
97. Renard, D.; Molenberghs, G.; Geys, H. A pairwise likelihood approach to estimation in multilevel probit models. *Comput. Stat. Data Anal.* **2004**, *44*, 649–667. doi:10.1016/S0167-9473(02)00263-3.

98. Varin, C.; Reid, N.; Firth, D. An overview of composite likelihood methods. *Stat. Sin.* **2011**, *21*, 5–42.
99. Feddag, M.L.; Bacci, S. Pairwise likelihood for the longitudinal mixed Rasch model. *Comput. Stat. Data Anal.* **2009**, *53*, 1027–1037. doi:10.1016/j.csda.2008.08.031.
100. Andrich, D. Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika* **2010**, *75*, 292–308. doi:10.1007/S11336-010-9154-8.
101. Christensen, K.B. A Multidimensional Latent Class Rasch Model for the Assessment of the Health-Related Quality of Life. In *Rasch Models in Health*; Christensen, K.B., Kreiner, S., Mesbah, M., Eds.; Wiley: Hoboken, NJ, USA, 2013; pp. 49–62. doi:10.1002/9781118574454.ch3.
102. Draxler, C.; Tutz, G.; Zink, K.; Güter, C. Comparison of maximum likelihood with conditional pairwise likelihood estimation of person parameters in the Rasch model. *Commun. Stat. Simul. Comput.* **2016**, *45*, 2007–2017. doi:10.1080/03610918.2014.889153.
103. van der Linden, W.J.; Eggen, T.J.H.M. An empirical Bayesian approach to item banking. *Appl. Psychol. Meas.* **1986**, *10*, 345–354. doi:10.1177/014662168601000403.
104. Zwinderman, A.H. Pairwise parameter estimation in Rasch models. *Appl. Psychol. Meas.* **1995**, *19*, 369–375. doi:10.1177/014662169501900406.
105. de Gruijter, D.N. On the robustness of the “minimum-chi-square” method for the Rasch model. *Tijdschr Onderwijsres* **1987**, *12*, 225–232.
106. Fischer, G.H. *Einführung in Die Theorie Psychologischer Tests [Introduction to the Theory of Psychological Testing]*; Huber: Bern, Switzerland, 1974.
107. Choppin, B. A fully conditional estimation procedure for Rasch model parameters. *Eval. Educ.* **1982**, *9*, 29–42.
108. Heine, J.H.; Tarnai, C. Pairwise Rasch model item parameter recovery under sparse data conditions. *Psych. Test Assess. Model.* **2015**, *57*, 3–36.
109. Wang, J.; Engelhard, G. A pairwise algorithm in R for rater-mediated assessments. *Rasch Meas. Trans.* **2014**, *28*, 1457–1459.
110. Finch, H.; French, B.F. A comparison of estimation techniques for IRT models with small samples. *Appl. Meas. Educ.* **2019**, *32*, 77–96. doi:10.1080/08957347.2019.1577243.
111. Garner, M. An eigenvector method for estimating item parameters of the dichotomous and polytomous Rasch models. *J. Appl. Meas.* **2002**, *3*, 107–128.
112. Garner, M.; Engelhard, G. Using paired comparison matrices to estimate parameters of the partial credit Rasch measurement model for rater-mediated assessments. *J. Appl. Meas.* **2009**, *10*, 30–41.
113. Saaty, T.L.; Vargas, L.G. Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios. *Math. Model.* **1984**, *5*, 309–324. doi:10.1016/0270-0255(84)90008-3.
114. Saaty, T.L.; Vargas, L.G. *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*; Springer: Berlin/Heidelberg, Germany, 2012. doi:10.1007/978-1-4614-3597-6.
115. Anderson, C.J.; Böckenholt, U. Graphical regression models for polytomous variables. *Psychometrika* **2000**, *65*, 497–509. doi:10.1007/BF02296340.
116. Anderson, C.J.; Vermunt, J.K. Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociol. Methodol.* **2000**, *30*, 81–121. doi:10.1111/0081-1750.00076.
117. Anderson, C.J.; Yu, H.T. Log-multiplicative association models as item response models. *Psychometrika* **2007**, *72*, 5–23. doi:10.1007/s11336-005-1419-2.
118. Anderson, C.J.; Li, Z.; Vermunt, J.K. Estimation of models in a Rasch family for polytomous items and multiple latent variables. *J. Stat. Softw.* **2007**, *20*, 1–36. doi:10.18637/jss.v020.i06.
119. Holland, P.W. The Dutch identity: A new tool for the study of item response models. *Psychometrika* **1990**, *55*, 5–18. doi:10.1007/BF02294739.
120. Wolfinger, R.; O’connell, M. Generalized linear mixed models a pseudo-likelihood approach. *J. Stat. Comput. Simul.* **1993**, *48*, 233–243. doi:10.1080/00949659308811554.
121. Le, L.T.; Adams, R.J. *Accuracy of Rasch Model Item Parameter Estimation*; ACER: Camberwel, UK, 2013.
122. Kim, Y.; Choi, Y.K.; Emery, S. Logistic regression with multiple random effects: A simulation study of estimation methods and statistical packages. *Am. Stat.* **2013**, *67*, 171–182. doi:10.1080/00031305.2013.817357.
123. Robitzsch, A. A comparison of estimation methods for the Rasch model. In *Book of Short Papers—SIS 2021*; Perna, C., Salvati, N., Spagnolo, F.S., Eds.; Pearson: Upper Saddle River, NJ, USA, 2021; pp. 157–162.
124. R Core Team. *R: A Language and Environment for Statistical Computing*; Vienna, Austria, 2020. Available online: <https://www.R-project.org/> (accessed on 20 August 2020).
125. Robitzsch, A.; Steinfeld, J. *immer: Item Response Models for Multiple Ratings*; R Package Version 1.1-35. 2018. Available online: <https://CRAN.R-project.org/package=immer> (accessed on 10 December 2020).
126. Heine, J.H. *pairwise: Rasch Model Parameters by Pairwise Algorithm*; R Package Version 0.5.0-2. 2021. Available online: <https://CRAN.R-project.org/package=pairwise> (accessed on 6 January 2021).
127. Li, Z.; Hong, F. *plRasch: Log Linear by Linear Association Models and Rasch Family Models by Pseudolikelihood Estimation*; R Package Version 1.0. 2014. Available online: <https://CRAN.R-project.org/package=plRasch> (accessed on 10 January 2014).
128. Robitzsch, A. *Sirt: Supplementary Item Response Theory Models*; R Package Version 3.10-111. 2021. Available online: <https://github.com/alexanderrobitzsch/sirt> (accessed on 25 June 2021).

129. Robitzsch, A.; Lüdtke, O. Reflections on analytical choices in the scaling model for test scores in international large-scale assessment studies. *PsyArXiv* **2021**, doi:10.31234/osf.io/pkjth.
130. Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*; Lord, F.M., Novick, M.R., Eds.; MIT Press: Reading, MA, USA, 1968; pp. 397–479.
131. Finch, H. Estimation of item response theory parameters in the presence of missing data. *J. Educ. Meas.* **2008**, *45*, 225–245, doi:10.1111/j.1745-3984.2008.00062.x.
132. Mislevy, R.J. Missing responses in item response modeling. In *Handbook of Item Response Theory, Vol. 2: Statistical Tools*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 171–194, doi:10.1201/b19166-10.
133. Waterbury, G.T. Missing data and the Rasch model: The effects of missing data mechanisms on item parameter estimation. *J. Appl. Meas.* **2019**, *20*, 154–166.
134. Kubinger, K.D.; Steinfeld, J.; Reif, M.; Yanagida, T. Biased (conditional) parameter estimation of a Rasch model calibrated item pool administered according to a branched testing design. *Psych. Test Assess. Model.* **2012**, *54*, 450–460.
135. Eggen, T.J.H.M.; Verhelst, N.D. Item calibration in incomplete testing designs. *Psicológica* **2011**, *32*, 107–132.
136. Fox, J.P. *Bayesian Item Response Modeling*; Springer: New York, NY, USA, 2010; doi:10.1007/978-1-4419-0742-4.
137. Hadfield, J.D. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J. Stat. Softw.* **2010**, *33*, 1–22, doi:10.18637/jss.v033.i02.
138. Kim, S.H.; Cohen, A.S.; Kwak, M.; Lee, J. Priors in Bayesian estimation under the Rasch model. *J. Appl. Meas.* **2019**, *20*, 384–398.
139. Luo, Y.; Jiao, H. Using the Stan program for Bayesian item response theory. *Educ. Psychol. Meas.* **2018**, *78*, 384–408, doi:10.1177/0013164417693666.
140. Rupp, A.A.; Dey, D.K.; Zumbo, B.D. To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Struct. Equ. Model.* **2004**, *11*, 424–451, doi:10.1207/s15328007sem1103_7.
141. Swaminathan, H.; Gifford, J.A. Bayesian estimation in the Rasch model. *J. Educ. Stat.* **1982**, *7*, 175–191, doi:10.3102/10769986007003175.
142. Draxler, C. Bayesian conditional inference for Rasch models. *AStA Adv. Stat. Anal.* **2018**, *102*, 245–262, doi:10.1007/s10182-017-0303-6.
143. Huang, P.H. Penalized least squares for structural equation modeling with ordinal responses. *Multivar. Behav. Res.* **2020**, doi:10.1080/00273171.2020.1820309.