NEW ITERATIVE ALGORITHMS FOR ESTIMATION OF ITEM FUNCTIONING

A PREPRINT

Adéla Hladká[1,2], **Patrícia Martinková**[1,3], **Marek Brabec**[1]
[1] Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic
[2] Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
[3] Faculty of Education, Charles University, Prague, Czech Republic

This is the authors' version of the manuscript.

ABSTRACT

When the item functioning of multi-item measurement is modeled with three or four-parameter models, parameter estimation may become challenging. Effective algorithms are crucial in such scenarios. This paper explores innovations to parameter estimation in generalized logistic regression models, which may be used in item response modeling to account for guessing/pretending or slipping/dissimulation and for the effect of covariates. We introduce a new implementation of the EM algorithm and propose a new algorithm based on the parametrized link function. The two novel iterative algorithms are compared to existing methods in a simulation study. Additionally, the study examines software implementation, including the specification of initial values for numerical algorithms and asymptotic properties with an estimation of standard errors. Overall, the newly proposed algorithm based on the parametrized link function outperforms other procedures, especially for small sample sizes. Moreover, the newly implemented EM algorithm provides additional information regarding respondents' inclination to guess or pretend and slip or dissimulate when answering the item. The study also discusses applications of the methods in the context of the detection of differential item functioning. Methods are demonstrated using real data from psychological and educational assessments.

1 Introduction

In psychometrics, Three-Parameter Logistic (3PL) and Four-Parameter Logistic (4PL) models are flexible tools that allow capturing complex item response patterns and accommodating a more comprehensive range of item characteristics, including possible guessing and slipping rates in the context of educational measurement and pretending and dissimulation in the context of psychological and health measurement. However, estimation in these models, both in the Item Response Theory (IRT) (Birnbaum, 1968; Barton & Lord, 1981) and non-IRT framework (Drabinová & Martinková, 2017; Hladká & Martinková, 2020), may become challenging due to several factors, including the complexity of these models caused by their non-linearity, high-dimensionality of the parameter space, and the nature of the data being analyzed. These models typically require a large sample size (Kim & Oshima, 2013), which can result in computationally demanding fitting. Therefore, efficient algorithms, advanced estimation techniques, and software implementation are crucial for the effectiveness and accessibility of these models' use in practice.

Recent research renewed interest in the 3–4PL IRT models since the availability of computing resources is on the rise. New approaches in estimation are being studied extensively (Battauz, 2020; Culpepper, 2016; Loken & Rulison, 2010; Meng, Xu, Zhang, & Tao, 2020; Fu, Zhang, Su, Shi, & Tao, 2021), which help solve some of the computational issues. However, their focus is mainly limited to large-scale assessments, while estimating item parameters with moderate sample sizes is still unreachable. To address the computational issues more effectively and accurately recover item characteristics such as guessing and slipping, the IRT models may benefit from traditional item analysis and Generalized Linear and Nonlinear Models (GLNMs), their simpler score-based counterparts (Martinková & Hladká, 2023). The GLNMs can offer improved and more precise starting values for related IRT models and allow for statistical inference regarding item parameters while still being less computationally complex compared to IRT models.

GLNMs incorporates a class of generalized logistic regression models that are natural extensions of the logistic regression model to describe item functioning. Analogously to 3–4PL IRT models, generalized logistic regression may account for the possibility that an item can be correctly answered or endorsed without the necessary knowledge or trait, e.g., due to guessing or pretending. In this case, the logistic regression model is extended by including a parameter defining a lower asymptote of the probability curve, which may be larger than zero. Similarly, the model can consider the possibility that an item is incorrectly answered or opposed by a respondent with a high level of a particular trait due to issues such as inattention, lack of time, or dissimulation; this model includes an upper asymptote of the probability curve, which may be lower than one. These models can be seen as score-based counterparts to 3–4PL IRT models since they assume the same shape of the item response curve; however, in contrast to the class of latent variable models, this approach uses an observed estimate of the underlying latent trait.

Furthermore, logistic regression, its extensions, and their latent variable counterparts have become widely used for identifying between-group differences on item level when responding to multi-item measurements (Swaminathan & Rogers, 1990). The phenomenon, known as Differential Item Functioning (DIF), indicates whether responses to an item vary for respondents with the same level of an underlying latent trait but from different groups (e.g., defined by gender, age, or socio-economic status). In this vein, DIF detection is essential for a deeper understanding of group differences, assessing the effectiveness of various treatments, or uncovering potential unfairness in educational tests. It is identified as one of the crucial topics in measurement (AERA, APA, & NCME, 2014).

Beyond the logistic-shaped models, various psychometric and statistical methods have been proposed for an important task of DIF identification. Typically two branches of DIF detection methods are recognized – score-based techniques and IRT-based approaches, while both lines are still being studied extensively (Schneider, Strobl, Zeileis, and Debelak, 2022; Schauberger and Tutz, 2016; Paek and Fukuhara, 2015; Suh and Bolt, 2011; for an overview of classical approaches, see Magis, Béland, Tuerlinckx, and De Boeck, 2010).

The estimation in the logistic regression model is a straightforward procedure, but extending the parametric space by including additional parameters in this model makes it more statistically and computationally challenging and demanding and may result in convergence issues. This is even more present in IRT modeling, where latent ability is estimated together with item parameters. In this vein, GLNMs can be seen as a helpful alternative in describing item functioning and identifying DIF, accounting for possible guessing or inattention while also being accessible in practice.

Therefore, this article examines innovations in the item parameter estimation for the GLNMs in the context of DIF detection. As the main contribution, the work proposes novel iterative algorithms, examines their theoretical properties, and compares the newly proposed methods to existing ones in a simulation study. The use of estimation procedures is then exemplified on real data examples with an application to DIF detection, with the secondary goal of providing possibilities for more accurate DIF detection.

The rest of the manuscript is organized as follows: To begin, Section 2 introduces the GLNMs and its relationship to IRT framework, examining the estimation techniques. This section provides a detailed description of two existing methods for parameter estimation, the Nonlinear Least Squares (NLS) and the Maximum Likelihood (ML) method, and their application to fitting GLNMs. Furthermore, as an alternative to the direct implementation of the ML method, this study proposes a novel implementation of the Expectation-Maximisation (EM) algorithm and a new approach based on a Parametrized Link Function (PLF). Additionally, this section provides asymptotic properties of the estimates, an estimation of standard errors, and a software implementation, including a specification of starting values in iterative algorithms. Subsequently, Section 3 describes the design and results of the simulation study. To illustrate differences and challenges between the existing and newly proposed methods in practice and the context of DIF detection, this work provides two real data analyses in Section 4. Finally, Section 5 contains the discussion and concluding remarks.

2 Methodology

2.1 Generalized linear and nonlinear models for item functioning

GLNMs extend the logistic regression model by accounting for the possibility of guessing or inattention when answering an item. The *simple 4PL model* describes functioning of the item i, meaning the probability of endorsing item i by respondent p, by introducing four parameters:

$$\pi_{pi} = \mathrm{P}(Y_{pi} = 1|\theta_p) = c_i + (d_i - c_i)\,\frac{\exp(b_{i0} + b_{i1}\theta_p)}{1 + \exp(b_{i0} + b_{i1}\theta_p)}, \tag{1}$$

with θ_p being an observed trait of respondent p.

Parameter interpretation. All four parameters have an intuitive interpretation: The parameters c_i and d_i are the lower and upper asymptotes of the probability sigmoid function $\pi_{pi}(x)$ since

$$\lim_{x\to-\infty} \pi_{pi}(x) = c_i \quad \text{and} \quad \lim_{x\to\infty} \pi_{pi}(x) = d_i,$$

where $c_i \in [0,1], d_i \in [0,1]$ and $c_i < d_i$ if $b_{i1} > 0$ and $c_i > d_i$ otherwise. Evidently, with $c_i = 0$ and $d_i = 1$, this model recovers a standard logistic regression for item i.

In psychological and health-related assessments, the asymptotes c_i may represent pretending or simulation, and $1 - d_i$ may represent the probability of reluctance to admit difficulties due to social norms or dissimulation. In educational testing, parameter c_i can be interpreted as the probability that the respondents guessed the correct answer without possessing the necessary knowledge θ_p, also known as a pseudo-guessing parameter. On the other hand, $1 - d_i$ can be viewed as the probability that respondents were inattentive while their knowledge θ_p was sufficient (Hladká & Martinková, 2020), or a lapse-rate (Kingdom & Prins, 2016). Next, the parameter b_{i0} is an intercept parameter related to the difficulty of item i (or item popularity in psychological and health-related assessments), and parameter b_{i1} is linked to a slope of the sigmoid curve $\pi_{pi}(x)$, which is also called discrimination of the respective item.

Adding covariates, group-specific 4PL model. The simple model (1) can be further extended by incorporating additional respondents' characteristics. That is, instead of using a single variable θ_p to describe item functioning, a vector of covariates $\boldsymbol{X}_p = (1, X_{p1}, \dots, X_{pk})^\intercal$, $p = 1, \dots, n$, is involved, including the original observed trait and an intercept term. This process produces extra parameters $\boldsymbol{b}_i = (b_{i0}, \dots, b_{ik})^\intercal$. Beyond this, even asymptotes may depend on respondents' characteristics $\boldsymbol{Z}_p = (1, Z_{p1}, \dots, Z_{pj})^\intercal$, $p = 1, \dots, n$, which are not necessarily the same as $\boldsymbol{X}_p$. This general *covariate-specific 4PL model* is of form

$$\pi_{pi} = \mathrm{P}(Y_{pi} = 1|\boldsymbol{X}_p, \boldsymbol{Z}_p) = \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i + (\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i)\,\frac{\exp(\boldsymbol{X}_p^\intercal \boldsymbol{b}_i)}{1 + \exp(\boldsymbol{X}_p^\intercal \boldsymbol{b}_i)}, \tag{2}$$

where $\boldsymbol{c}_i = (c_{i0}, \dots, c_{ij})^\intercal$ and $\boldsymbol{d}_i = (d_{i0}, \dots, d_{ij})^\intercal$ are asymptote parameters for item i. Note that we typically assume $\boldsymbol{Z}_p$ being categorical rather than continuous variables describing respondents' characteristics to keep meaningful interpretation while requiring $0 \leq \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i < \boldsymbol{Z}_p^\intercal \boldsymbol{d}_i \leq 1$.

As a special case of the covariate-specific 4PL model (2), an additional single binary covariate G_p might be considered. This grouping variable describes a respondent's membership to a social group ($G_p = 0$ for the reference group and $G_p = 1$ for the focal group). In other words, this special case assumes $\boldsymbol{X}_p = (1, \theta_p, G_p, \theta_p \cdot G_p)^\intercal$ and $\boldsymbol{Z}_p = (1, G_p)^\intercal$, which reduces the covariate-specific 4PL model (2) to a group-specific form:

$$\begin{aligned}\pi_{pi} = \mathrm{P}(Y_{pi} = 1|\theta_p, G_p) &= c_i + c_{i\mathrm{DIF}} G_p \\ &+ (d_i - d_{i\mathrm{DIF}} G_p - c_i - c_{i\mathrm{DIF}} G_p)\,\frac{\exp(b_{i0} + b_{i1}\theta_p + b_{i2} G_p + b_{i3}\theta_p \cdot G_p)}{1 + \exp(b_{i0} + b_{i1}\theta_p + b_{i2} G_p + b_{i3}\theta_p \cdot G_p)}.\end{aligned} \tag{3}$$

Testing for DIF. The *group-specific 4PL model* (3) can be then used for testing between-group differences on the item level with a DIF analysis (Hladká & Martinková, 2020) with, for example, the likelihood-ratio test. The likelihood-ratio test measures the difference between the log-likelihood l_{i1} of the larger model (e.g., the group-specific 4PL model (3)) and the log-likelihood l_{i0} of its submodel (e.g., the simple 4PL model (1)) for the given item i. The resulting LR_i statistic has an asymptotic χ^2-distribution under the smaller model with degrees of freedom equal to a difference in the number of parameters in the two models:

$$LR_i = -2\,(l_{i0} - l_{i1}) \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2(\mathrm{df}_{i1} - \mathrm{df}_{i0}). \tag{4}$$

Similarly, any two nested submodels of the group-specific 4PL model (3) can be compared to test for the significance of group-related item parameters.

Matching criterion. In these models, θ_p is an observed variable describing the measured trait of the respondent, such as anxiety, fatigue, quality of life, or math ability, here called the *matching criterion*. In the context of the logistic regression method for DIF detection, the total test score (or its standardized version) is typically used as the matching criterion (Swaminathan & Rogers, 1990). Other options for the matching criterion include a pre-test score (to identify differential item functioning in change; see Martinková, Hladká, and Potužníková, 2020), a score on another test measuring the same construct, or an estimate of the latent trait provided by an IRT model.

IRT framework. Within the IRT framework, the matching criterion θ_p in the models (1)–(3) is replaced by a latent ability θ, necessitating joint estimation with item parameters. Both frameworks share the same shape of item characteristic curves with a comparable interpretation. A notable distinction lies in the estimation process: IRT models simultaneously estimate parameters for all items, which is typically not the case for models (1)–(3) as described below. However, GLNMs entail lower computational demands, as they require smaller sample sizes to yield precise estimates. The estimation algorithms for GLNMs, which are the focus of this paper, may thus be further incorporated into the IRT framework as is further described in the Discussion.

2.2 Estimation of item parameters

Numerous algorithms are available to estimate item parameters in the covariate-specific 4PL model (2). First, this section describes two methods that may be directly implemented in the existing software: The NLS method and the ML method. This study discusses the asymptotic properties of the estimates. Next, the study introduces two newly proposed iterative algorithms, which might improve the implementation of the computationally demanding ML method: The EM algorithm inspired by the work of Dinse (2011) and an iterative algorithm based on PLF.

2.2.1 Nonlinear least squares

The parameter estimates of the covariate-specific 4PL model (2) can be determined using the NLS method (Dennis, Gay, & Welsch, 1981; Drabinová & Martinková, 2017; Hladká & Martinková, 2020), which is based on minimisation of the Residual Sum of Squares (RSS) of item i with respect to item parameters $(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$:

$$\mathrm{RSS}_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i) = \sum_{p=1}^{n} \left[Y_{pi} - \pi_{pi}\right]^2 = \sum_{p=1}^{n} \left[Y_{pi} - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i - (\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{d}_i - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i)\frac{\exp(\boldsymbol{X}_p^\mathsf{T}\boldsymbol{b}_i)}{1+\exp(\boldsymbol{X}_p^\mathsf{T}\boldsymbol{b}_i)}\right]^2, \tag{5}$$

where n is the number of respondents. Since the criterion function $\mathrm{RSS}_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$ is continuously differentiable with respect to item parameters $(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$, the minimiser can be obtained when the gradient is zero. Thus, the minimization process involves a calculation of the first partial derivatives with respect to item parameters $(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$ and finding a solution of relevant nonlinear estimating equations (e.g., van der Vaart, 1998, Chapter 5). Since $\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i$ and $\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{d}_i$ asymptotes represent probabilities, it is necessary to ensure that these expressions are kept in the interval of $[0, 1]$ which is accomplished using numerical approaches.

The asymptotic properties of the NLS estimator, such as consistency and asymptotic distribution, can be derived under the classical set of regularity conditions (e.g., van der Vaart, 1998, Theorems 5.41 and 5.42; see also Appendix A.1 for more details). This study proposes a sandwich estimator (A1), which can be used as a natural estimate of the asymptotic variance of the NLS estimate.

2.2.2 Maximum likelihood

The second option for estimating item parameters in the covariate-specific 4PL model (2) is the ML method (Hladká & Martinková, 2020). Using a notation $\phi_{pi} = \frac{\exp(\boldsymbol{X}_p^\mathsf{T}\boldsymbol{b}_i)}{1+\exp(\boldsymbol{X}_p^\mathsf{T}\boldsymbol{b}_i)}$, the corresponding likelihood function for item i has the following form:

$$L_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i) = \prod_{p=1}^{n} \left[\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i + (\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{d}_i - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i)\phi_{pi}\right]^{Y_{pi}} \left[1 - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i - (\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{d}_i - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i)\phi_{pi}\right]^{1-Y_{pi}},$$

and the log-likelihood function is then given by

$$\begin{aligned} l_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i) = \sum_{p=1}^{n} \big\{&Y_{pi}\log(\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i + (\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{d}_i - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i)\phi_{pi}) \\ &+ (1-Y_{pi})\log(1 - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i - (\boldsymbol{Z}_p^\mathsf{T}\boldsymbol{d}_i - \boldsymbol{Z}_p^\mathsf{T}\boldsymbol{c}_i)\phi_{pi})\big\}. \end{aligned}$$

The parameter estimates are obtained by maximization of the log-likelihood function. Thus, this approach proceeds similarly to the logistic regression model, except for a larger dimension of the parametric space. To find the maximizer of the log-likelihood function $l_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$, the first partial derivatives are set to zero and these so-called likelihood equations must be solved. However, the solution of a system of nonlinear equations cannot be derived algebraically and needs to be numerically estimated using a suitable iterative process.

Using van der Vaart's (1998) Theorems 5.41 and 5.42, consistency and asymptotic normality can be shown for the ML estimator, see Appendix A.2 for more details. Additionally, the estimate of the asymptotic variance of the item parameters is an inverse of the observed information matrix (A2).

2.2.3 EM algorithm

The ML method may be computationally demanding, and iterative algorithms might help in those situations. Inspired by the work of Dinse (2011), this study adopts a version of the EM algorithm (Dempster, Laird, & Rubin, 1977) for parameter estimation in the covariate-specific 4PL model (2).

The original problem can be reformulated using latent variables that describe the hypothetical response status of respondents (see also Zheng, Meng, Guo, & Liu, 2018, for the 3PL model). In our setting, we consider four mutually exclusive latent variables (W_{pi1}, W_{pi2}, W_{pi3}, W_{pi4}), where variable $W_{pij} = 1$ indicates that respondent p belongs in the category $j = 1, \dots, 4$ for an item i, whereas $W_{pij} = 0$ indicates that respondent does not belong in this category.

In the context of educational, psychological, health-related, or other types of multi-item measurement, the four categories can be interpreted as follows: Categories 1 and 2 indicate whether a respondent who responded correctly to item i or endorsed it (i.e., $Y_{pi} = 1$) was determined to do so ($W_{pi1} = 1$, e.g., the respondent guessed correct answer while their knowledge or ability was insufficient, or the respondent simulated described situation) or not ($W_{pi2} = 1$, e.g., had sufficient knowledge or ability to answer correctly and did not guess, or endorsed while experiencing described situation). On the other hand, Categories 3 and 4 indicate whether the respondent who did not respond correctly or did not endorse the item (i.e., $Y_{pi} = 0$) was prone to do so ($W_{pi3} = 1$, e.g., did not have sufficient knowledge, ability, or trait) or not ($W_{pi4} = 1$, e.g., incorrectly answered, or did not endorse, due to another reason such as inattention, lack of time, or dissimulation). Thus, the observed indicator Y_{pi} and its complement $1 - Y_{pi}$ could be rewritten as $Y_{pi} = W_{pi1} + W_{pi2}$ and $1 - Y_{pi} = W_{pi3} + W_{pi4}$ (Figure 1). While the latent variables W_{pi2} and W_{pi3} represent desired response styles, W_{pi1} and W_{pi4} represent response styles which are undesired.

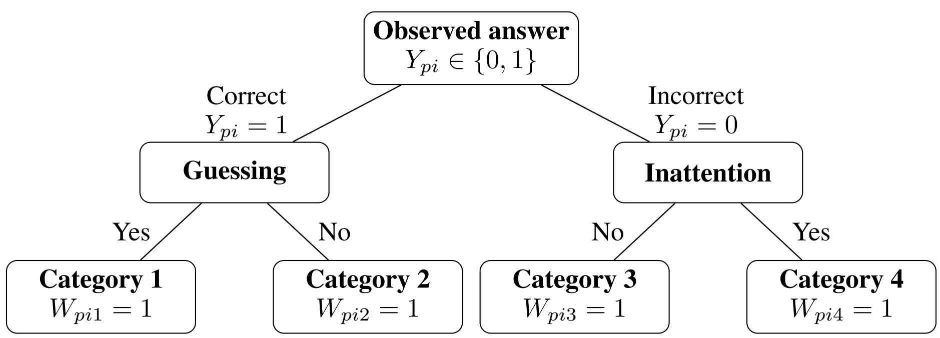


Figure 1: Graphical representation of the relationships among latent variables for the EM algorithm

Let $\boldsymbol{Z}_p^\intercal \boldsymbol{c}_i$ be the regressor-based probability that the respondent was determined to respond to item i correctly or endorse it (Category 1), and let $\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i$ be the regressor-based probability of the respondent not prone to respond correctly or endorse item i (Categories 1–3). Then $\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i$ gives the regressor-based probability that the respondent was not determined but prone to (Categories 2 and 3). Further, we denote ϕ_{pi} and $1 - \phi_{pi}$ – the probabilities to answer a given item correctly (Category 2) and incorrectly (Category 3), respectively, depending on the regressors $\boldsymbol{X}_p$. Finally, the probability that the respondent did not respond correctly and was not prone to do so is given by $1 - (\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i) - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i = 1 - \boldsymbol{Z}_p^\intercal \boldsymbol{d}_i$ (Category 4). In summary, the expected values of the latent variables are then given by the following terms

$$\boldsymbol{Z}_p^\intercal \boldsymbol{c}_i, \;\; (\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i)\phi_{pi}, \;\; (\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i)(1 - \phi_{pi}), \;\; 1 - \boldsymbol{Z}_p^\intercal \boldsymbol{d}_i,$$

and the probability of a correct response or endorsement is given by

$$\begin{aligned} \mathrm{P}(Y_{pi} = 1 | \boldsymbol{X}_p) &= \mathrm{P}(W_{pi1} + W_{pi2} = 1 | \boldsymbol{X}_p) = \mathrm{P}(W_{pi1} = 1 | \boldsymbol{X}_p) + \mathrm{P}(W_{pi2} = 1 | \boldsymbol{X}_p) \\ &= \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i + (\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i)\phi_{pi}, \end{aligned}$$

which under the logistic model $\phi_{pi} = \frac{\exp(\boldsymbol{X}_p^\intercal \boldsymbol{b}_i)}{1+\exp(\boldsymbol{X}_p^\intercal \boldsymbol{b}_i)}$ produces the covariate-specific 4PL model (2).

Using the setting of the latent variables, the corresponding log-likelihood function for item i takes the following form:

$$\begin{aligned} l_i^{\text{EM}} =& \sum_{p=1}^{n} \left[W_{pi2} \log\left(\phi_{pi}\right) + W_{pi3} \log\left(1 - \phi_{pi}\right)\right] \\ &+ \sum_{p=1}^{n} \left[W_{pi1} \log\left(\boldsymbol{Z}_p^\intercal \boldsymbol{c}_i\right) + W_{pi4} \log\left(1 - \boldsymbol{Z}_p^\intercal \boldsymbol{d}_i\right) + \ (W_{pi2} + W_{pi3}) \log\left(\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i\right)\right] \\ =& \, l_{i1}^{\text{EM}} + l_{i2}^{\text{EM}}. \end{aligned}$$

The log-likelihood function l_{i1}^{EM} includes only parameters $\boldsymbol{b}_i$ and regressors $\boldsymbol{X}_p$, whereas the log-likelihood function l_{i2}^{EM} incorporates only parameters related to the asymptotes of the sigmoid function and includes only regressors $\boldsymbol{Z}_p$. Notably, the log-likelihood function l_{i1}^{EM} has a form of the log-likelihood function for the logistic regression. However, in contrast to the logistic regression model, in this setting, it does not necessarily hold that $W_{pi2} + W_{pi3} = 1$ since the correct answer could be guessed or the respondent could be inattentive, producing $W_{pi2} + W_{pi3} = 0$. The log-likelihood function l_{i2}^{EM} takes the form of the log-likelihood for multinomial data with one trial and with the regressor-based probabilities $\boldsymbol{Z}_p^\intercal \boldsymbol{c}_i$, $\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i$, and $1 - \boldsymbol{Z}_p^\intercal \boldsymbol{d}_i$.

The EM algorithm estimates item parameters in two steps – expectation and maximization. These two steps are repeated until the convergence criterion is met, such as until the change in log-likelihood is lower than a predefined value.

Expectation. At the E-step, conditionally on the item responses Y_{pi} and the current parameter estimate $(\widehat{\boldsymbol{b}}_i, \widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{d}}_i)$, the estimates of latent variables are calculated as their expected values:

$$\begin{aligned} \widehat{W}_{pi1} &= \frac{\boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i Y_{pi}}{\boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i + (\boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{d}}_i - \boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i)\widehat{\phi}_{pi}}, & \widehat{W}_{pi2} &= Y_{pi} - \widehat{W}_{pi1}, \\ \widehat{W}_{pi4} &= \frac{\left(1 - \boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{d}}_i\right)(1 - Y_{pi})}{1 - \boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i - (\boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{d}}_i - \boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i)\widehat{\phi}_{pi}}, & \widehat{W}_{pi3} &= 1 - Y_{pi} - \widehat{W}_{pi4}. \end{aligned} \tag{6}$$

Maximization. At the M-step, conditionally on the current estimates of the latent variables $\widehat{W}_{pi2}$ and $\widehat{W}_{pi3}$, the estimates of parameters $\boldsymbol{b}_i$ maximize the log-likelihood function l_{i1}^{EM}. The estimates $\widehat{\boldsymbol{c}}_i$ and $\widehat{\boldsymbol{d}}_i$ are given by a maximisation of the log-likelihood function l_{i2}^{EM} conditionally on current estimates of the latent variables $\widehat{W}_{pi1}$, $\widehat{W}_{pi2}$, $\widehat{W}_{pi3}$, and $\widehat{W}_{pi4}$.

The EM algorithm is designed to gain the ML estimates of the item parameters, so estimates have the same asymptotic properties as described above.

Additionally, it might be of practical interest that the EM algorithm provides estimates of latent variables W_{pi1}, W_{pi2}, W_{pi3}, and W_{pi4}. Their mean values over all items may be interpreted in an educational context as follows: The $\overline{W}_{p1}$ as (undesired) "inclination to guess"; $\overline{W}_{p2}$ as a (desired) probability of "knowing correct answers when correctly answering"; $\overline{W}_{p3}$ as a (desired) probability of "not knowing correct answers when incorrectly answering"; and $\overline{W}_{p4}$ as (undesired) "inclination to slip/inattention". In the context of psychological or health-related measurements, the mean values may be interpreted as follows: The $\overline{W}_{p1}$ as the (undesired) "inclination to simulate"; $\overline{W}_{p2}$ as the (desired) probability of "endorsing while experiencing described situations"; $\overline{W}_{p3}$ as the (desired) probability of "not endorsing while not experiencing described situations"; and $\overline{W}_{p4}$ as the (undesired) "inclination to dissimulate".

2.2.4 Parametrized link function

In our setting, the covariate-specific 4PL model (2) can be viewed as a generalized linear model with a known PLF

$$g(\mu_{pi}; \boldsymbol{c}_i, \boldsymbol{d}_i) = \log\left(\frac{\mu_{pi} - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i}{\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \mu_{pi}}\right), \tag{7}$$

where the parameters $\boldsymbol{c}_i$ and $\boldsymbol{d}_i$ are unknown and may depend on regressors $\boldsymbol{Z}_p$. Subsequently, the mean function is determined by $\mu_{pi} = \pi_{pi}$ as given by (2) with a linear predictor $\boldsymbol{X}_p^\intercal \boldsymbol{b}_i$.

Keeping this setting in mind, this study proposes a new two-stage algorithm to estimate item parameters using the PLF (7), which involves repeating two steps until the convergence criterion is fulfilled.

Step one. First, conditionally on current estimates $\widehat{\boldsymbol{c}}_i$ and $\widehat{\boldsymbol{d}}_i$ of the PLF, the estimates of parameters $\boldsymbol{b}_i$ maximise the following log-likelihood function:

$$l_{i1}^{\text{PL}}(\boldsymbol{b}_i|\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{d}}_i) = \sum_{p=1}^{n} \Big\{ Y_{pi} \log(\boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i + (\boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{d}}_i - \boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i)\phi_{pi}) \\ + (1 - Y_{pi}) \log(1 - \boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i - (\boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{d}}_i - \boldsymbol{Z}_p^\intercal \widehat{\boldsymbol{c}}_i)\phi_{pi}) \Big\}.$$

The log-likelihood function $l_{i1}^{\text{PL}}(\boldsymbol{b}_i|\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{d}}_i)$ has a similar form to the log-likelihood function $l_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$ using the ML method. However, the parameters $\boldsymbol{c}_i$ and $\boldsymbol{d}_i$ are here replaced by their current estimates, $\widehat{\boldsymbol{c}}_i$ and $\widehat{\boldsymbol{d}}_i$.

Step two. Next, estimates $\widehat{\boldsymbol{c}}_i$ and $\widehat{\boldsymbol{d}}_i$ of the PLF (7) are calculated conditionally on the current estimates $\widehat{\boldsymbol{b}}_i$ as the arguments of the maxima of the following log-likelihood function

$$l_{i2}^{\text{PL}}(\boldsymbol{c}_i, \boldsymbol{d}_i|\widehat{\boldsymbol{b}}_i) = \sum_{p=1}^{n} \Big\{ Y_{pi} \log(\boldsymbol{Z}_p^\intercal \boldsymbol{c}_i + (\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i)\widehat{\phi}_{pi}) \\ + (1 - Y_{pi}) \log(1 - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i - (\boldsymbol{Z}_p^\intercal \boldsymbol{d}_i - \boldsymbol{Z}_p^\intercal \boldsymbol{c}_i)\widehat{\phi}_{pi}) \Big\}.$$

Again, the parameters $\boldsymbol{b}_i$ are replaced by their estimates $\widehat{\boldsymbol{b}}_i$, and ϕ_{pi} is thus replaced by $\widehat{\phi}_{pi}$.

In summary, the division into the two sets of parameters makes the algorithm based on PLF easy to implement in the `R` software and can take advantage of its existing functions. Because the algorithm is designed to produce the ML estimates, their asymptotic properties are the same as described above.

2.3 Implementation and software

For all analyses, software `R`, version 4.3.1 (R Core Team, 2023) was used. The NLS method was implemented using the base `nls()` function and the `"port"` algorithm (Gay, n.d.). The sandwich estimator (A1) of the asymptotic covariance matrix was computed using the **calculus** package (Guidotti, 2022). The ML estimation was performed with the base `optim()` function and the `"L-BFGS-B"` algorithm (Byrd, Lu, Nocedal, & Zhu, 1995). The EM algorithm implements directly (6) in the expectation step using the base `glm()` function and the `multinom()` function from the **nnet** package (Venables & Ripley, 2002) in the maximization step. Next, step one of the newly proposed algorithm based on PLF is implemented with the base `glm()` function with the modified logit link, which includes asymptote parameters. The asymptote parameters are estimated in step two using the base `optim()` function. The maximum number of iterations was 2,000 for all four methods, and the convergence criterion was set to 10^{-6} when possible.

Initial values. Starting values for item parameters were calculated as follows: The respondents were divided into three groups based upon tertiles of the matching criterion θ_p. Next, the asymptote parameters were estimated: c was computed as an empirical probability for those whose matching criterion was smaller than its average value in the first group defined by tertiles. The asymptote d was calculated as an empirical probability of those whose matching criterion was greater than its average value in the last group defined by tertiles. The slope parameter b_1 was estimated as a difference between the mean empirical probabilities of the last and the first group multiplied by 4. This difference is sometimes called the upper-lower index. Finally, the intercept b_0 was calculated as follows: First, a center point between the asymptotes was computed, and then we looked for the level of the matching criterion that would have corresponded to this empirical probability. Additionally, smoothing and corrections for the variability of the matching criterion were applied.

3 Simulation study

A simulation study was performed to compare various procedures to estimate parameters in the generalized logistic regression model, including the NLS, the ML method, the EM algorithm, and the newly proposed algorithm based on PLF. Two models were considered – the simple 4PL model (1) and the group-specific 4PL model (3).

3.1 Simulation design

Data generation. To generate data, ten sets with different combinations of item parameters were considered. The item parameters were chosen to correspond to common values: Parameters b_0, b_2, and b_3 were generated from the standard normal distribution, parameter b_1 was generated from a normal distribution with a mean value equal to 2.5, and a standard deviation of 0.5. Parameter c was generated from uniform distribution $\mathcal{U}(0.05, 0.30)$ for both groups. Parameter d was generated from uniform distribution $\mathcal{U}(0.7, 0.95)$ for both groups. In the case of the simple 4PL model (1), only parameters b_0, b_1, c, and d were considered. Next, the matching criterion θ_p was generated from the standard normal distribution for all respondents. Since item parameters are estimated item by item in generalized logistic models, we focused on items separately. We used a generated variable as the matching criterion instead of the total score. Binary responses were generated from the Bernoulli distribution with the calculated probabilities based upon the chosen 4PL model, true parameters, and the matching criterion variable. The sample size was set to $n =$ 500; 1,000; 2,500; and 5,000, i.e., 250; 500; 1,250; and 2,500 per group in the case of the group-specific 4PL model (3). Each scenario was replicated 1,000 times.

Simulation evaluation. To compare estimation methods, we first computed mean and median numbers of iteration runs and the convergence status of the methods, meaning the percentage of converged simulation runs; the percentage of runs that crashed (caused an error when fitting, e.g., due to singularities); and the percentage of those which reached the maximum number of iterations without convergence. Next, we selected only those simulation runs for which all four estimation methods converged successfully. We computed the mean parameter estimates and parametric confidence intervals, i.e., average intervals found for estimated standard errors derived for the respective algorithm. When confidence intervals for asymptote parameters exceeded their boundaries of 0 or 1, confidence intervals were truncated at the boundary value. The proportion of confidence intervals covering the true parameter value was calculated. Subsequently, the mean bias (i.e., the mean difference between estimates and true values) and Root Mean Squared Error (RMSE) (i.e., the square root of the average of squared errors) were calculated with respect to sample size. Finally, for a deeper insight into ML-based methods (i.e., traditionally implemented ML, the EM algorithm, and the algorithm based on PLF), we compared log-likelihoods for these three methods to those based on true values of parameters.

3.2 Simulation results

Convergence status. All four methods had low percentages of simulation runs that crashed for all sample sizes in the simple 4PL model (1). Still, the rate was mildly increased in the group-specific 4PL model (3) for the NLS method (6.72%) and for the algorithm based on PLF (9.22%) when $n = 500$. With the increasing sample size, convergence issues disappeared. The EM algorithm struggled to converge in a predefined number of iterations, especially for small sample sizes in both models. Additionally, the method based on PLF reached the maximum limit of 2,000 iterations only in a small percentage of simulation runs when smaller sample sizes were considered (Table 1).

Number of iterations. Furthermore, the methods differed in the number of iterations needed until the estimation process successfully ended. The EM algorithm yielded the largest mean and median numbers of iterations, which were somehow overestimated by simulation runs that did not finish without convergence (i.e., the maximum limit of 2,000 iterations was reached). The fewest iterations were needed for the NLS method. As expected, all the procedures required fewer simulation runs when the simple 4PL model (1) was considered than in the group-specific 4PL model (3). Beyond this, the number of iterations was decreasing with the increasing sample size in both models for all the methods (Table 1).

In both models, some estimation procedures produced non-meaningful estimates of parameters b_0–b_3 (absolute value over 100) despite successful convergence. Such simulations affected mean values significantly, so they were removed from a computation of the mean estimates and their confidence intervals for all four estimation methods. Incidence was similar for all methods (Table 1). Such estimates might be obtained due to insufficient sample size or starting values far from the global maximizer.

Parameter estimates. In the simple 4PL model (1), the PLF-based algorithm gained the most precise estimates of parameters b_0 and b_1 (in the sense of bias and RMSE) when smaller sample sizes were considered ($n = 500$ or $n = 1,000$). Additionally, the NLS method yielded slightly more biased estimates in these scenarios. The precision of the estimation improved for both parameters when the sample size increased in all four methods, whereas differences between estimation procedures narrowed. The accuracy of the estimates of the asymptote parameters c and d was similar for all four methods. The NLS method yielded the least biased estimates, while the PLF-based algorithm produced the lowest RMSE. However, the differences between estimation approaches were minor. The proportion of confidence intervals covering true values of item parameters was high for all four methods (Table 2). Slightly higher coverage for the NLS method was caused by somewhat larger confidence intervals.

In the group-specific 4PL model (3), the PLF-based algorithm yielded the most precise estimates of parameters b_0–b_3 in the sense of RMSE, especially for the smaller sample sizes. On the other hand, the NLS method produced the

Table 1: Convergence status, proportion of suspicious simulation runs, and the number of iterations for the four estimation methods

Method	Simple 4PL model (1)						Group-specific 4PL model (3)					
	Convergence status [%]				Number of iterations		Convergence status [%]				Number of iterations	
	Conv.	Crash.	DNF	Susp.	Mean	Median	Conv.	Crash.	DNF	Susp.	Mean	Median
						n = 500						
NLS	99.48	0.52	0.00	0.06	10.18	8.00	93.28	6.72	0.00	0.50	17.27	14.00
MLE	99.84	0.16	0.00	0.06	23.84	23.00	98.87	1.13	0.00	0.45	123.75	104.00
EM	93.70	0.00	6.30	0.03	347.14	152.00	93.38	0.07	6.55	0.50	452.89	211.00
PLF	98.34	1.02	0.64	0.03	144.31	18.00	89.84	9.22	0.94	0.47	248.06	66.00
						n = 1,000						
NLS	99.93	0.07	0.00	0.00	7.66	7.00	97.90	2.10	0.00	0.15	12.50	10.00
MLE	99.89	0.11	0.00	0.00	22.24	22.00	99.99	0.01	0.00	0.06	106.85	99.00
EM	95.76	0.00	4.24	0.00	295.53	149.00	93.50	0.00	6.50	0.10	447.87	201.00
PLF	99.71	0.11	0.18	0.00	97.73	14.00	97.84	1.96	0.20	0.10	172.43	47.00
						n = 2,500						
NLS	99.98	0.02	0.00	0.00	5.97	6.00	99.39	0.61	0.00	0.01	8.17	7.00
MLE	99.94	0.06	0.00	0.00	20.98	20.00	100.00	0.00	0.00	0.00	94.19	91.00
EM	96.23	0.00	3.77	0.00	254.47	134.00	94.55	0.00	5.45	0.01	356.17	161.00
PLF	99.98	0.00	0.02	0.00	51.77	9.00	99.87	0.12	0.01	0.01	95.36	25.00
						n = 5,000						
NLS	100.00	0.00	0.00	0.00	5.26	5.00	99.97	0.03	0.00	0.00	6.57	6.00
MLE	99.95	0.05	0.00	0.00	20.50	20.00	100.00	0.00	0.00	0.00	90.21	89.00
EM	97.64	0.00	2.36	0.00	223.45	127.00	95.90	0.00	4.10	0.01	308.32	141.00
PLF	100.00	0.00	0.00	0.00	24.23	8.00	99.98	0.02	0.00	0.02	50.03	11.00

Note. Conv. = converged, Crash. = crashed, DNF = did not finish, Susp. = suspicious.

largest RMSE in such scenarios. Computed bias was similar for all four methods. Similar to the simple 4PL model (1), the differences in the precision of the parameter estimates were narrowed with the increasing sample size, and all four estimation approaches gave estimates close to the true values of the item parameters. The estimates of the asymptote parameters c, c_{DIF}, d, and d_{DIF} were similar for all four methods. The EM algorithm provided slightly less biased mean estimates of the asymptote parameters, while the PLF-based algorithm produced slightly smaller RMSE. The proportion of confidence intervals covering true values of item parameters was high and similar for all four methods (Table 3). Different lengths of computed intervals caused differences in coverage of true parameters between the estimation methods.

Log-likelihood comparison. In the simple 4PL model (1), the algorithm based on PLF yielded log-likelihood values nearest to those computed based on true parameters in 91.31% of cases, followed by the EM algorithm in 8.45% and the directly implemented ML method in 0.23% of cases. There were similar differences between the three ML-based methods in the group-specific 4PL model (3). The algorithm based on PLF outperformed other likelihood-based estimation procedures in 87.56% of cases, while the EM algorithm worked the best in 12.06% and the ML method in 0.38%.

4 Real data examples

4.1 Data description

We demonstrate the estimation procedures with an application to DIF detection on two real-data examples – item responses on a questionnaire on PROMIS anxiety scale[1] and a test measuring learning competence (Martinková et al., 2020; Martinková & Drabinová, 2018).

[1]http://www.nihpromis.org

Table 2: Bias, RMSE, and confidence interval (CI) coverage by four estimation methods using the simple model (1) with respect to sample size

Method	Bias				RMSE				CI coverage [%]
	500	1,000	2,500	5,000	500	1,000	2,500	5,000	
b_0									
NLS	0.044	0.012	0.003	0.003	1.028	0.399	0.208	0.143	96.27
MLE	0.045	0.016	0.004	0.004	0.902	0.387	0.204	0.141	95.89
EM	0.035	0.015	0.003	0.003	0.839	0.388	0.204	0.141	95.75
PLF	0.016	0.010	0.006	0.006	0.514	0.322	0.196	0.144	94.95
b_1									
NLS	−0.645	−0.217	−0.070	−0.037	2.876	0.990	0.446	0.300	95.44
MLE	−0.507	−0.183	−0.059	−0.031	2.253	0.932	0.433	0.292	95.60
EM	−0.470	−0.174	−0.055	−0.026	2.194	0.922	0.430	0.291	95.47
PLF	−0.159	−0.033	0.024	0.026	1.011	0.628	0.377	0.278	95.19
c									
NLS	0.003	0.003	0.001	0.000	0.061	0.044	0.027	0.019	94.62
MLE	0.006	0.004	0.001	0.001	0.063	0.044	0.027	0.019	94.34
EM	0.006	0.004	0.002	0.001	0.062	0.043	0.027	0.019	94.27
PLF	0.012	0.009	0.005	0.004	0.060	0.042	0.027	0.019	95.16
d									
NLS	−0.001	−0.001	0.000	0.000	0.065	0.049	0.032	0.023	94.28
MLE	−0.003	−0.002	−0.000	0.000	0.065	0.049	0.031	0.022	93.88
EM	−0.003	−0.001	−0.000	0.000	0.065	0.049	0.031	0.022	93.77
PLF	−0.009	−0.006	−0.003	−0.002	0.062	0.047	0.030	0.022	94.12

4.1.1 Anxiety scale

The Anxiety dataset consisted of responses to 29 Likert-type questions (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Always) from 766 respondents. Additionally, the dataset included information on the respondents' age, education, and gender (0 = Male and 1 = Female). Overall, there were 369 male participants and 397 female participants.

For this work, item responses were dichotomized as follows: 0 was assigned to response Never (i.e., response $= 1$ on the original scale), while 1 was given to responses Rarely and more often (i.e., response ≥ 2 on the original scale).

4.1.2 Learning competence

The LearningToLearn dataset consisted of binary-coded responses from 782 subjects to (mostly) multiple-choice test consisting of 41 items within seven subscales. Each respondent was tested twice – the first time in the 6th grade and the second time in the 9th grade; responses from the 6th grade only were considered for this analysis. Among other variables, the dataset included information on the school track of respondents (basic school track = 0, academic school track = 1). Overall, 391 students attended basic school, and 391 pursued selective academic school.

4.2 Real data analysis design

This work considered the simple 4PL model (1) and the group-specific 4PL model (3) with different constraints on asymptote parameters. In the Anxiety dataset, the lower asymptotes were set to zeros, i.e., $c_i = 0$ and $c_{i\text{DIF}} = 0$, since pretending (i.e., lower asymptote greater than 0) was not expected. On the other hand, in the LearningToLearn dataset, the upper asymptotes were set to ones, i.e., $d_i = 1$ and $d_{i\text{DIF}} = 0$, since slipping (i.e., upper asymptote lower than 1) was not expected.

The matching criterion θ_p in the Anxiety dataset was the overall level of anxiety, which was calculated as a standardized sum of non-dichotomized item responses. Similarly, the standardized total test score gained in the 6th grade was used

Table 3: Bias, RMSE, and confidence interval (CI) coverage by four estimation methods using the group-specific model (3) with respect to sample size

Method	Bias				RMSE				CI
	500	1,000	2,500	5,000	500	1,000	2,500	5,000	coverage [%]
b_0									
NLS	−0.025	0.008	0.007	0.003	2.230	1.259	0.351	0.204	96.63
MLE	0.002	0.023	0.009	0.004	1.777	1.105	0.341	0.200	96.07
EM	0.023	0.026	0.008	0.005	1.770	1.086	0.337	0.200	95.40
PLF	0.009	0.006	0.010	0.011	0.846	0.485	0.285	0.232	94.71
b_1									
NLS	−1.400	−0.640	−0.156	−0.072	6.027	3.041	0.896	0.445	94.64
MLE	−1.013	−0.506	−0.122	−0.059	4.745	2.463	0.768	0.431	95.37
EM	−0.811	−0.475	−0.116	−0.053	4.289	2.371	0.768	0.427	94.12
PLF	−0.214	−0.120	0.010	0.036	1.510	1.000	0.541	0.390	93.67
b_2									
NLS	0.082	0.016	−0.001	0.002	3.005	1.528	0.477	0.289	96.91
MLE	0.006	−0.012	−0.007	−0.001	2.467	1.357	0.463	0.284	96.34
EM	−0.022	−0.011	−0.006	0.000	2.270	1.277	0.461	0.283	95.59
PLF	0.024	0.005	−0.012	−0.013	1.319	0.763	0.401	0.309	94.53
b_3									
NLS	0.047	0.031	−0.000	−0.003	8.105	4.068	1.175	0.649	97.73
MLE	−0.113	0.023	−0.003	−0.004	6.367	3.366	1.040	0.630	97.46
EM	0.098	0.061	0.003	0.001	5.889	3.046	1.040	0.625	95.93
PLF	0.084	0.101	0.065	0.050	2.329	1.532	0.800	0.584	95.06
c									
NLS	0.015	0.005	0.003	0.001	0.087	0.062	0.039	0.027	94.25
MLE	0.020	0.008	0.004	0.001	0.091	0.063	0.040	0.027	94.23
EM	0.024	0.008	0.004	0.001	0.091	0.063	0.039	0.027	93.33
PLF	0.029	0.015	0.009	0.006	0.089	0.061	0.039	0.027	94.51
c_{DIF}									
NLS	−0.007	−0.002	−0.001	−0.001	0.113	0.083	0.053	0.037	95.19
MLE	−0.007	−0.002	−0.001	−0.001	0.116	0.085	0.053	0.037	95.13
EM	−0.004	−0.002	−0.001	−0.001	0.117	0.084	0.052	0.037	93.63
PLF	−0.005	−0.000	0.001	0.001	0.110	0.080	0.052	0.037	94.58
d									
NLS	−0.010	−0.003	−0.003	−0.001	0.084	0.065	0.044	0.031	93.72
MLE	−0.015	−0.005	−0.004	−0.001	0.087	0.066	0.045	0.031	93.46
EM	−0.018	−0.005	−0.003	−0.001	0.088	0.065	0.043	0.031	91.56
PLF	−0.022	−0.012	−0.008	−0.005	0.084	0.063	0.043	0.031	92.31
d_{DIF}									
NLS	0.003	0.002	0.001	0.000	0.118	0.091	0.059	0.043	95.29
MLE	0.003	0.002	0.002	0.000	0.121	0.092	0.060	0.043	95.17
EM	−0.000	0.001	0.001	0.000	0.121	0.091	0.058	0.042	92.77
PLF	0.000	−0.000	−0.001	−0.001	0.114	0.087	0.058	0.042	93.57

as the matching criterion θ_p in the LearningToLearn dataset. For the group-specific models, the grouping variable G_p

was defined by the gender of respondents for the Anxiety dataset and the school track for the LearningToLearn dataset. DIF detection was performed concerning these variables.

The two newly proposed estimation methods were applied for the two datasets and models: the EM algorithm and the algorithm based on PLF. The same approach for computing starting values as in the simulation study was used to analyze both datasets. In the case of convergence issues, the initial values were re-calculated based on successfully converged estimates provided by other methods.

In this study, item parameter estimates were computed and reported with confidence intervals. Next, the likelihood-ratio test (4) was performed to compare the two nested models (simple and group-specific) to identify the DIF for all items and both novel estimation methods. Finally, for the EM algorithm, mean values of estimated latent variables over all items were computed. A significance level of 0.05 was used for all the tests.

4.3 Real data analysis results

4.3.1 Anxiety scale

DIF detection. Using the likelihood-ratio test, the simple 4PL model (1) with constraints on lower asymptotes was rejected for items R6 (*"I was concerned about my mental health"*; p-value = 0.001 considering either estimation algorithm), R7 (*"I felt upset"*; p-value = 0.038), R10 (*"I had sudden feelings of panic"*; p-value = 0.031), R21 (*"I had twitching or trembling muscles"*; p-value = 0.035), and R29 (*"I had difficulty calming down"*; p-value = 0.04) by using either of the two newly proposed estimation algorithms (i.e., these items functioned differently). In these items, the less restrictive group-specific model (3) was preferred, allowing for different intercepts, slopes, and upper asymptotes for the two groups (i.e., these items functioned differently).

We now take a closer look at DIF item R7, for which the confidence interval of estimated parameter d_{DIF} (i.e., the difference in upper asymptote between groups of female and male respondents) did not cover 0 (Table A2). According to the model, the male respondents with high anxiety levels seemed not to admit feeling upset with a probability of $1 - 0.92 = 0.08$, while there was no dissimulation rate for female respondents (Figure 2).

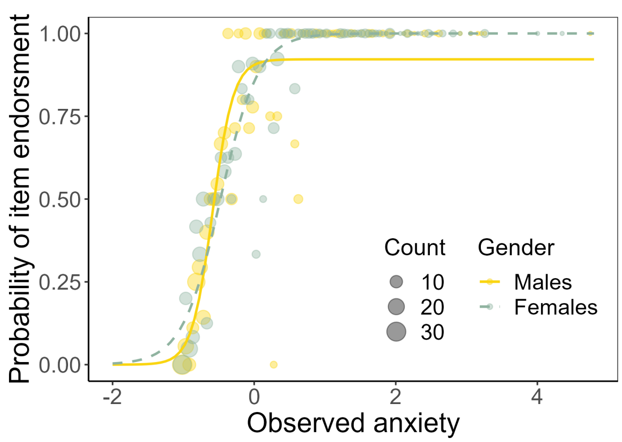


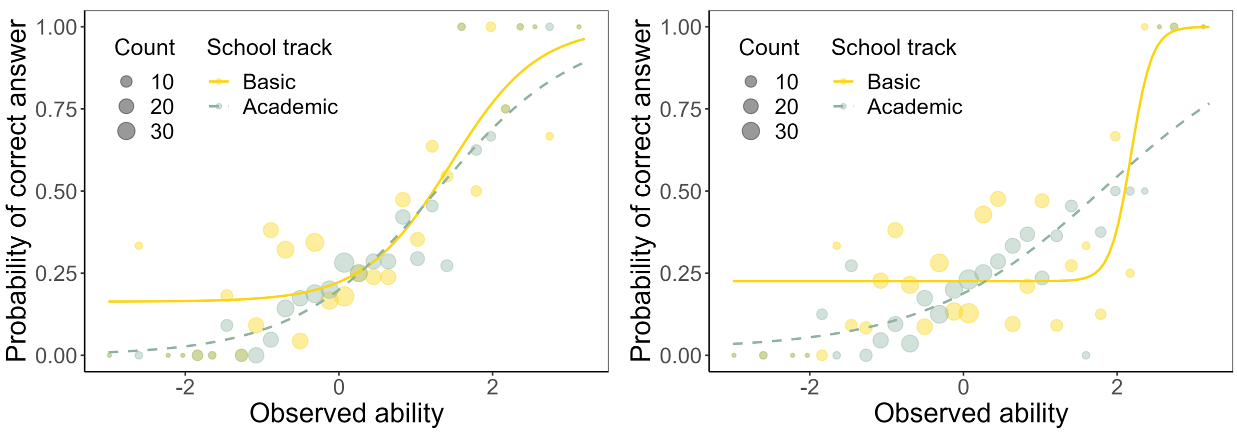
Figure 2: Estimated item characteristic curves of item R7 of the Anxiety dataset for the group-specific 4PL model (3) with constraints on lower asymptotes

Latent variables estimates by EM algorithm. Using the group-specific model (3), male respondent 272 with an overall level of anxiety equal to 0.62 had the highest "inclination to dissimulate," equal to 0.20, meaning they would have dissimulated almost six items out of the 29-item Anxiety dataset. On the other hand, female respondent 264 with the same overall anxiety level had a probability equal to 0.06, which would correspond to the dissimulation of less than two items.

4.3.2 Learning competence

DIF detection. Using the likelihood-ratio test, the simple 4PL model (1) with constraints on upper asymptotes was rejected for items 1A (p-value = 0.006 considering either estimation algorithm), 1D (p-value = 0.009), 6F (p-value = 0.043), 6H (p-value = 0.032), and 7F (p-value = 0.049) by using either of the two newly proposed estimation algorithms. In these items, the less restrictive group-specific model (3) was preferred, allowing for different intercepts, slopes, and lower asymptotes for the two groups.

Items 6F and 6H were identified as functioning differently due to differences in lower asymptotes, i.e., confidence interval of estimated parameter c_{DIF} (i.e., the difference in lower asymptotes between the two school tracks) did not cover 0 (Table A4). In both items, students from the basic school track tended to guess more often than students from the academic school track. In item 6F, the probability of guessing in the basic school track was 0.16, while in the academic school track, it was 0 (Figure 3a). In item 6H, the probability of guessing in the basic school track was 0.23, while in the academic school track, it was 0.02 (Figure 3b).



(a) Item 6F

(b) Item 6H

Figure 3: Estimated item characteristic curves of selected DIF items of the LearningToLearn dataset for the group-specific 4PL model (3) with constraints on upper asymptotes

Both items were related to ”solving tasks with invented mathematical operators which are conditionally defined depending on the value of the digits they connect” (Martinková et al., 2020). The original study suggested that students from academic schools might have been trying to solve these difficult items more often, while students from the basic school track might have been guessing more often.

Latent variables estimates by EM algorithm. Considering the group-specific model (3), respondent 486, who attended basic school with an overall level of learning competence of -0.70, has the highest ”inclination to guess,” equal to 0.38, meaning they would guess almost 16 items out of the 41-item test on learning competencies. On the other hand, respondent 386, who attended academic school and who had exactly the same level of learning competence, has a probability equal to 0.17, corresponding to the guessing of 7 items out of 41.

5 Discussion

This work explored novel approaches for estimating item functioning within the GLNMs framework. The study proposed two iterative procedures (a procedure using the EM algorithm and a new method based on PLF) as alternatives to the directly implemented ML method. The methods were compared via simulation with existing algorithms and implemented in `R`.

In the simulation study, the traditional NLS approach produced the most biased parameter estimates with wide confidence intervals. The directly implemented ML method performed satisfactorily; however, the newly proposed methods were superior in some aspects: The EM algorithm provided slightly less biased parameter estimates than the directly implemented ML method, and it more often produced log-likelihood values were closer to those computed based on true parameters. These were at the price of a higher number of iterations being needed for this approach to converge, while the maximum number of iterations was reached in several cases. As an added value, the EM algorithm provided additional information on respondents' latent response styles. The newly proposed algorithm based on PLF yielded the least biased parameter estimates of the expit function for most settings, especially when small sample sizes and additional covariates were considered. Moreover, in most scenarios, the PLF-based algorithm yielded log-likelihood values nearest to those computed based on true underlying parameters. Conversely, there was a higher rate of crashed simulations for the group-specific 4PL model (3) and small sample size. The precision of the asymptote parameters was similar for all four estimation techniques. As the sample size increased, differences between the estimation methods vanished, and all estimates were near the true values of the item parameters.

Using two real data examples, we illustrated the possible benefits of generalized logistic regression models in item response modeling, estimating asymptotes, and their application to DIF analysis. Further, we presented how the practitioners may benefit from the added value of the EM algorithm, which can be used to estimate the probability of guessing correctly answered items (in the context of psychological assessment, endorsing an item due to pretending) or answering incorrectly due to inattention (in the context of psychological assessment, not endorsing an item due to dissimulation) for individual respondents. We also demonstrated practical challenges in estimation procedures, including specifying initial values.

While the EM algorithm is a well-established estimation procedure, we use it in the previously not-considered context of GLNMs in multi-item measurements, and we further extend the original approach to a group-specific model and more covariates. On the other hand, the PLF-based algorithm is novel and has not been proposed in this form for parameter estimation in the generalized logistic regression model. However, in recent decades, the idea of the PLF has been extensively discussed in the literature by many authors in various contexts, including Basu and Rathouz (2005), Flach (2014), and Scallan, Gilchrist, and Green (1984). For example, Pregibon (1980) proposed the ML estimation of the link parameters using a weighted least squares algorithm. Similarly, McCullagh and Nelder (1989) adapted this approach and presented an algorithm in which several models with the fixed link functions were fitted. Furthermore, Kaiser (1997) proposed a modified scoring algorithm to perform simultaneous ML estimation of all parameters. Scallan et al. (1984) proposed an iterative two-stage algorithm, building on the work of Richards (1961). This study examined generalized logistic regression, accounting for the possibility of guessing/pretending and inattention/dissimulation, whereas these features may depend upon the respondents' characteristics.

The crucial part of each estimation process is specifying starting values for item parameters because these values may significantly impact the speed and precision of the estimation process. For instance, initial values far from the true item parameters may lead to situations where the estimation algorithm returns only a local extreme or does not converge. In this work, we used an approach based on an upper-lower index, resulting in low convergence rate issues with satisfactory estimation precision. However, other possible naive estimates of discrimination (and other parameters) could be considered, such as a correlation between an item score and the total test score without a given item.

This study has several limitations, and several possible further directions exist. First, the simulation study was limited to two models – the simple 4PL model (1) and the group-specific 4PL model (3), both of which included only one or two covariates. The simulation study suggested requiring a larger sample size with an increasing number of covariates. Second, this article described the NLS method as a simple approach, not accounting for the heteroscedasticity of binary data. For such data, Pearson's residuals might be more appropriate to use. This weighted form (e.g., Ritz, Baty, Streibig, & Gerhard, 2015) takes the original squares of residuals and divides them by the variance $\pi_{pi}(1-\pi_{pi})$. Next, the RSS of item i (5) would take the following form:

$$\mathrm{RSS}_i(\boldsymbol{\gamma}_i) = \sum_{p=1}^{n} \frac{\left(Y_{pi} - \pi_{pi}\right)^2}{\pi_{pi}\left(1 - \pi_{pi}\right)}.$$

However, the number of observations on the tails of the matching criterion is typically tiny and provides only small variability at most. These heavy weights would require a nearly exact fit for cases with few observations. Nevertheless, the computation of the NLS estimates demonstrated in this work was straightforward and efficient, providing sufficient precision. Thus, this method could be helpful in some instances, such as producing an initial idea about parameter values and using these estimates as starting values for other approaches. Third, it is important to acknowledge that the estimation methods studied here can be sensitive to the choice of optimization algorithm and the control parameters. The directly implemented ML estimation was performed with the "L-BFGS-B" algorithm to account for constraints in asymptotes. Alternatively, asymptote parameters may depend on covariates through a transformation function, so the estimating algorithm does not need to incorporate constraints. The performance of these two approaches might differ. Moreover, the control parameters were set the same for all estimation methods, while the sensitivity of methods to their setting may vary and may be imposed in different quantities in different algorithms (e.g., deviation, likelihood, or the norm of gradient vector)

While the primary focus of this paper lies in enhancing parameter estimation within GLNMs for multi-item measurement, it also touches upon the application of these algorithms in DIF detection. We illustrated the DIF detection by comparing the largest and the smallest models; however, a step-by-step procedure omitting the parameters with non-significant effects might be applied in practice to explore DIF in detail. Although DIF detection is not the central theme, practical examples illustrate the significance of assessing the fairness and validity of assessments across diverse groups. Nevertheless, this study does not aim to evaluate the properties of the underlying DIF detection procedure or to compare it with popular existing methods such as the anchor item-based approaches (Candell & Drasgow, 1988; Clauser, Mazor, & Hambleton, 1993; W.-C. Wang & Yeh, 2003; Kopf, Zeileis, & Strobl, 2015) or more recent regularization based approaches (Magis, Tuerlinckx, & De Boeck, 2015; Tutz & Schauberger, 2015; Belzak & Bauer, 2020; C. Wang, Zhu, & Xu, 2023).

Establishing a common scale on which respondents from different groups can be scored and ranked is a crucial step in DIF analysis. In both the IRT and non-IRT frameworks (including, e.g., the Mantel-Haenszel test or SIBTEST procedure), the inclusion of DIF items in estimation or computation of ability estimate may have a severe impact on which items are detected as functioning differently. One possibility for dealing with such an issue is applying an item purification iterative algorithm (Lord, 1980; see also Hladká, Martinková, & Magis, 2024).

In contrast to the IRT framework, GLNMs offer flexibility in selecting the matching criterion. Besides the standardized total scores, the model may utilize latent trait estimates - possibly in an iterative algorithm, yielding an IRT model. The model may also utilize previous test scores as a matching criterion, allowing to study the differential item functioning in change (Martinková et al., 2020), for further applications, also see (Kolek, Šisler, Martinková, & Brom, 2021; Kolek, Martinková, Vařejková, Šisler, & Brom, 2024), or other relevant criterion variables. Additionally, the covariate-specific model (2) accommodates multi-dimensional matching criteria, similar to its IRT counterpart. Both frameworks share the same objective when accounting for the same underlying latent trait – to estimate item functioning with a logistic-shaped item characteristic curve. In such instances, the estimating algorithms for GLNMs can provide initial estimates for the corresponding IRT model, as they are less computationally demanding, requiring lower sample size and resulting in fewer convergence issues. Moreover, they may be used for the iterative estimation of ability and item parameters.

This study's real data examples explored item functioning in the multi-item measurement related to anxiety and learning competencies. However, the parameter estimation task in the presented models would also be relevant to many other educational, psychological, and health-related measurement areas, such as the assessment of well-being, fatigue, reading literacy, and others. Moreover, the generalized logistic regression model is not limited to multi-item measurements since the class determined by Equation (2) represents a broad family of the covariate-specific 4PL models. This model might be used and further extended in various study fields, including but not limited to quantitative pharmacology (Dinse, 2011), applied microbiology (Brands, Schulze Struchtrup, Stamminger, & Bockmühl, 2020), modeling patterns of urban electricity usage (To, Lai, Lo, Lam, & Chung, 2012), and plant growth modeling (Zub, Rambaud, Béthencourt, & Brancourt-Hulmel, 2012). Therefore, the estimation procedures proposed in this work are highly relevant for a wide range of researchers and practitioners, both within and outside the psychometric field.

To conclude, this study researched advances in fitting generalized logistic regression models using various estimation techniques, including two newly proposed ones. We demonstrated the superiority of the novel implementation of the EM algorithm and the newly proposed method based on PLF over the existing NLS and directly implemented ML methods. Improving estimation algorithms is critical since it could increase precision while maintaining a user-friendly implementation. It may also provide additional information regarding individual respondents and items; thus, it is worth investing resources in the advancements of estimation methods.

Acknowledgement

The study was funded by the Czech Science Foundation project "Theoretical Foundations of Computational Psychometrics" grant number 21-03658S, by the project "Research of Excellence on Digital Technologies and Wellbeing CZ.02.01.01/00/22_008/0004583" which is co-financed by the European Union, and by the institutional support RVO 67985807.

Online Supplementary Material

Accompanying R scripts are available at https://osf.io/eu5zm/.

References

American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *ETS Research Report Series*, *1981*(1), 1–8. doi: 10.1002/j.2333-8504.1981.tb01255.x

Basu, A., & Rathouz, P. J. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics*, *6*(1), 93–109. doi: 10.1093/biostatistics/kxh020

Battauz, M. (2020). Regularized estimation of the four-parameter logistic model. *Psych*, *2*(4), 269–278. doi: 10.3390/psych2040020

Belzak, W., & Bauer, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychological Methods*, *25*(6), 673. doi: 10.1037/met0000253

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Addison-Wesley, Reading, MA.

Brands, B., Schulze Struchtrup, S., Stamminger, R., & Bockmühl, D. P. (2020). A method to evaluate factors influencing the microbial reduction in domestic dishwashers. *Journal of Applied Microbiology*, *128*(5), 1324–1338. doi: 10.1111/jam.14564

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208. doi: 10.1137/0916069

Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement*, *12*(3), 253–260. doi: 10.1177/014662168801200304

Clauser, B., Mazor, K., & Hambleton, R. K. (1993). The effects of purification of matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education*, *6*(4), 269–279. doi: 10.1207/s15324818ame0604_2

Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*, *81*(4), 1142–1163. doi: 10.1007/s11336-015-9477-6

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x

Dennis, J. E. J., Gay, D. M., & Welsch, R. E. (1981). An adaptive nonlinear least-squares algorithm. *Transactions on Mathematical Software*, *7*(3), 348–368. doi: 10.1145/355958.355965

Dinse, G. E. (2011). An EM algorithm for fitting a four-parameter logistic model to binary dose-response data. *Journal of Agricultural, Biological, and Environmental Statistics*, *16*(2), 221–232. doi: 10.1007/s13253-010-0045-3

Drabinová, A., & Martinková, P. (2017). Detection of differential item functioning with nonlinear regression: A non-IRT approach accounting for guessing. *Journal of Educational Measurement*, *54*(4), 498–517. doi: 10.1111/jedm.12158

Flach, N. (2014). *Generalized linear models with parametric link families in R* (Unpublished master's thesis). Technische Universität München, Department of Mathematics, München.

Fu, Z., Zhang, S., Su, Y.-H., Shi, N., & Tao, J. (2021). A Gibbs sampler for the multidimensional four-parameter logistic item response model via a data augmentation scheme. *British Journal of Mathematical and Statistical Psychology*, *74*(3), 427–464. doi: 10.1111/bmsp.12234

Gay, D. (n.d.). *Port library documentation.* `http://www.netlib.org/port/`. (Accessed: 2022-12-13)

Guidotti, E. (2022). calculus: High-dimensional numerical and symbolic calculus in R. *Journal of Statistical Software*, *104*(5), 1–37. doi: 10.18637/jss.v104.i05

Hladká, A., & Martinková, P. (2020). difNLR: Generalized logistic regression models for DIF and DDF detection. *The R Journal*, *12*(1), 300–323. doi: 10.32614/RJ-2020-014

Hladká, A., Martinková, P., & Magis, D. (2024). Combining item purification and multiple comparison adjustment methods in detection of differential item functioning. *Multivariate Behavioral Research*, *59*(1), 46–61. doi: 10.1080/00273171.2023.2205393

Hogg, R. V., McKean, J., & Craig, A. T. (2018). *Introduction to mathematical statistics* (Eighth ed.). Pearson Education.

Kaiser, M. S. (1997). Maximum likelihood estimation of link function parameters. *Computational Statistics & Data Analysis*, *24*(1), 79–87. doi: 10.1016/S0167-9473(96)00055-2

Kim, J., & Oshima, T. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement*, *73*(3), 458–470. doi: 10.1177/0013164412467033

Kingdom, F. A., & Prins, N. (2016). *Psychophysics: A practical introduction* (Second ed.). Academic Press. doi: 10.1016/C2012-0-01278-1

Kolek, L., Martinková, P., Vařejková, M., Šisler, V., & Brom, C. (2024). Is video games' effect on attitudes universal? examining the effects of perspective-taking game mechanics and attitude importance. *Journal of Computer Assisted Learning*, *40*(2), 667–684.

Kolek, L., Šisler, V., Martinková, P., & Brom, C. (2021). Can video games change attitudes towards history? results from a laboratory experiment measuring short- and long-term effects. *Journal of Computer Assisted Learning*, *37*(5), 1348–1369. doi: 10.1111/jcal.12575

Kopf, J., Zeileis, A., & Strobl, C. (2015). Anchor selection strategies for DIF analysis: Review, assessment, and new approaches. *Educational and Psychological Measurement*, *75*(1), 22–56. doi: 10.1177/0013164414529792

Loken, E., & Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 509–525. doi: 10.1348/000711009X474502

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Milton Park, Abingdon, Oxon: Routledge. doi: 10.4324/9780203056615

Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, *42*(3), 847–862. doi: 10.3758/BRM.42.3.847

Magis, D., Tuerlinckx, F., & De Boeck, P. (2015). Detection of differential item functioning using the lasso approach. *Journal of Educational and Behavioral Statistics*, *40*(2), 111–135. doi: 10.3102/1076998614559747

Martinková, P., & Drabinová, A. (2018). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal*, *10*(2), 503–515. doi: 10.32614/RJ-2018-074

Martinková, P., & Hladká, A. (2023). *Computational aspects of psychometric methods: With R*. Chapman and Hall/CRC. doi: 10.1201/9781003054313

Martinková, P., Hladká, A., & Potužníková, E. (2020). Is academic tracking related to gains in learning competence? Using propensity score matching and differential item change functioning analysis for better understanding of tracking implications. *Learning and Instruction*, *66*, 101286. doi: 10.1016/j.learninstruc.2019.101286

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (Second ed.). Chapman & Hall.

Meng, X., Xu, G., Zhang, J., & Tao, J. (2020). Marginalized maximum a posteriori estimation for the four-parameter logistic model under a mixture modelling framework. *British Journal of Mathematical and Statistical Psychology*, *73*, 51–82. doi: 10.1111/bmsp.12185

Paek, I., & Fukuhara, H. (2015). An investigation of DIF mechanisms in the context of differential testlet effects. *British Journal of Mathematical and Statistical Psychology*, *68*(1), 142–157. doi: 10.1111/bmsp.12039

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29*(1), 15–24. doi: 10.2307/2346405

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Richards, F. S. (1961). A method of maximum-likelihood estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *23*(2), 469–475. doi: 10.1111/j.2517-6161.1961.tb00430.x

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-response analysis using R. *PLOS ONE*, *10*(12), e0146021. doi: 10.1371/journal.pone.0146021

Scallan, A., Gilchrist, R., & Green, M. (1984). Fitting parametric link functions in generalised linear models. *Computational Statistics & Data Analysis*, *2*(1), 37–49. doi: 10.1016/0167-9473(84)90031-8

Schauberger, G., & Tutz, G. (2016). Detection of differential item functioning in Rasch models by boosting techniques. *British Journal of Mathematical and Statistical Psychology*, *69*(1), 80–103. doi: 10.1111/bmsp.12060

Schneider, L., Strobl, C., Zeileis, A., & Debelak, R. (2022). An R toolbox for score-based measurement invariance tests in IRT models. *Behavior Research Methods*, *54*, 2101–2113. doi: 10.3758/s13428-021-01689-0

Suh, Y., & Bolt, D. M. (2011). A nested logit approach for investigating distractors as causes of differential item functioning. *Journal of Educational Measurement*, *48*(2), 188–205. doi: 10.1111/j.1745-3984.2011.00139.x

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*(4), 361–370. doi: 10.1111/j.1745-3984.1990.tb00754.x

To, W., Lai, T., Lo, W., Lam, K., & Chung, W. (2012). The growth pattern and fuel life cycle analysis of the electricity consumption of Hong Kong. *Environmental Pollution*, *165*, 1–10. doi: 10.1016/j.envpol.2012.02.007

Tutz, G., & Schauberger, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika*, *80*, 21–43.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press. doi: 10.1017/CBO9780511802256

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth ed.). New York: Springer. doi: 10.1007/978-0-387-21706-2

Wang, C., Zhu, R., & Xu, G. (2023). Using lasso and adaptive lasso to identify DIF in multidimensional 2PL models. *Multivariate Behavioral Research*, *58*(2), 387–407. doi: 10.1080/00273171.2021.1985950

Wang, W.-C., & Yeh, Y.-L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*, *27*(6), 479–498. doi: 10.1177/0146621603259902

Zheng, C., Meng, X., Guo, S., & Liu, Z. (2018). Expectation-maximization-maximization: A feasible MLE algorithm for the three-parameter logistic model based on a mixture modeling reformulation. *Frontiers in Psychology*, *8*, 2302. doi: 10.3389/fpsyg.2017.02302

Zub, H., Rambaud, C., Béthencourt, L., & Brancourt-Hulmel, M. (2012). Late emergence and rapid growth maximize the plant development of Miscanthus clones. *BioEnergy Research*, *5*(4), 841–854. doi: 10.1007/s12155-012-9194-2

Appendices

A Asymptotics

A.1 Nonlinear least squares

Asymptotic properties of the NLS estimator, such as consistency and asymptotic distribution, can be derived under the classical set of regularity conditions (e.g., van der Vaart, 1998, Theorems 5.41 and 5.42). Next, one must reformulate these conditions for the covariate-specific 4PL model (2) considering $\boldsymbol{\gamma}_{iX} = (\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$ is a vector of true parameters:

[R0] A vector of true parameters $\boldsymbol{\gamma}_{iX}$ satisfies

$$\mathsf{E}\left(\boldsymbol{\psi}_i(Y_{pi}, \boldsymbol{X}_p; \boldsymbol{\gamma}_{iX})\right) = \mathsf{E}\left(-2(Y_{pi} - \pi_{pi})\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_{iX}}\right) = \mathbf{0}.$$

[R1] The true parameter $\boldsymbol{\gamma}_{iX}$ is an interior point of the parameter space.

[R2] The function $\boldsymbol{\psi}_i(y, \boldsymbol{x}; \boldsymbol{\gamma}_i)$ is twice continuously differentiable with respect to $\boldsymbol{\gamma}_i$ for every $(y, \boldsymbol{x})$.

[R3] For each $\boldsymbol{\gamma}_i^*$ in a neighbourhood of $\boldsymbol{\gamma}_{iX}$, there exists an integrable function $\ddot{\boldsymbol{\psi}}(y, \boldsymbol{x})$ such that

$$\left|\frac{\partial^2 \psi_{ik}(y, \boldsymbol{x}; \boldsymbol{\gamma}_i)}{\partial \gamma_{ij} \partial \gamma_{il}}\right| \leq \ddot{\boldsymbol{\psi}}(y, \boldsymbol{x}),\ \forall k, j, l.$$

[R4] The matrix

$$\begin{aligned}\mathbb{\Gamma}_i(\boldsymbol{\gamma}_i) &= \mathsf{E}\left(\frac{\partial \boldsymbol{\psi}_i(Y_{pi}, \boldsymbol{X}_p; \boldsymbol{\gamma}_i)}{\partial \boldsymbol{\gamma}_i^\top}\right)\\ &= 2\,\mathsf{E}\left(\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\left(\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\right)^\top - (Y_{pi} - \pi_{pi})\frac{\partial^2 \pi_{pi}}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^\top}\right)\\ &= 2\,\mathsf{E}\left(\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\left(\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\right)^\top\right)\end{aligned}$$

is finite and regular in a neighborhood of $\boldsymbol{\gamma}_{iX}$.

[R5] The variance matrix

$$\begin{aligned}\mathbb{\Sigma}_i(\boldsymbol{\gamma}_i) &= \mathsf{E}\left(\boldsymbol{\psi}_i(Y_{pi}, \boldsymbol{X}_p; \boldsymbol{\gamma}_i)\boldsymbol{\psi}_i^\top(Y_{pi}, \boldsymbol{X}_p; \boldsymbol{\gamma}_i)\right)\\ &= 4\,\mathsf{E}\left((Y_{pi} - \pi_{pi})^2\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\left(\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\right)^\top\right)\\ &= 4\,\mathsf{E}\left(\pi_{pi}(1 - \pi_{pi})\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\left(\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}\right)^\top\right)\end{aligned}$$

is finite for $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_{iX}$.

Specifically, Theorem 5.42 (van der Vaart, 1998, p. 68) implies that under the conditions [R0]–[R5], the probability that the estimating equations, $\frac{\partial \mathrm{RSS}_i}{\partial \gamma_{ik}} = 0$, have at least one root tends to 1, as $n \to \infty$, and there exists a sequence $\hat{\boldsymbol{\gamma}}_i$ (depending on n) such that $\hat{\boldsymbol{\gamma}}_i \xrightarrow[n\to\infty]{P} \boldsymbol{\gamma}_{iX}$. Moreover, the sequence $\hat{\boldsymbol{\gamma}}_i$ can be chosen as a local maximum for each n. Theorem 5.41 (van der Vaart, 1998, p. 68) demonstrates that every consistent estimator $\hat{\boldsymbol{\gamma}}_i$ has asymptotically normal distribution, that is:

$$\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{iX}\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \mathbb{\Gamma}_i^{-1}(\boldsymbol{\gamma}_{iX})\mathbb{\Sigma}_i(\boldsymbol{\gamma}_{iX})\mathbb{\Gamma}_i^{-1}(\boldsymbol{\gamma}_{iX})\right),$$

Clearly, the conditions [R0] and [R2] hold. To satisfy the condition [R1], one must bound asymptote parameters $\boldsymbol{c}_i$ and $\boldsymbol{d}_i$ to open intervals. Suppose the asymptote parameters are on the boundary of the parameter space (e.g., $c_i = 0$, $d_i = 1$, and $c_{i\mathrm{DIF}} = d_{i\mathrm{DIF}} = 0$ in the group-specific 4PL model (3)). In that case, the logistic regression model may be used instead. Additionally, the model (3) with some of the parameters fixed (e.g., $d_i = 1$ and $d_{i\text{DIF}} = 0$), may be considered analogously. Notably, the asymptotic properties derived here will hold for these submodels. However, it is impossible to test whether the full model or its submodel better fits the data using, for instance, the likelihood-ratio test. Regarding the condition [R3], in this case, a polynomial of x of the fourth degree can be taken as an integrable dominating function.

Furthermore, because $\boldsymbol{X}_p$ describes respondent characteristics such as the standardized total score or gender, one can assume that their range is bounded, so the partial derivatives $\frac{\partial \pi_{pi}}{\partial \boldsymbol{\gamma}_i}$ are also bounded. Thus, matrices $\mathbb{\Gamma}_i(\boldsymbol{\gamma}_i)$ and $\mathbb{\Sigma}_i(\boldsymbol{\gamma}_i)$ are both finite, and the condition [R5] holds. Finally, when the rows or columns of the matrix $\mathbb{\Gamma}_i(\boldsymbol{\gamma}_i)$ are linearly independent, the matrix has a full rank and thus is regular, satisfying condition [R4]. For instance, the singularity of the matrix may occur when $G_p = 0, \forall p$ (or $G_p = 1, \forall p$), meaning that all respondents are from the reference (or focal) group.

Therefore, all the assumptions [R1]–[R5] hold under the mild additional conditions, so $\hat{\boldsymbol{\gamma}}_i$ has desired asymptotic properties such as consistency and asymptotic normality of the item parameter estimates.

Estimate of asymptotic variance. The natural estimate of the asymptotic variance of $\hat{\boldsymbol{\gamma}}_i$ is a *sandwich estimator* given by

$$\frac{1}{n}\widehat{\mathbb{\Gamma}}_{in}^{-1}(\hat{\boldsymbol{\gamma}}_i)\widehat{\mathbb{\Sigma}}_{in}(\hat{\boldsymbol{\gamma}}_i)\widehat{\mathbb{\Gamma}}_{in}^{-1}(\hat{\boldsymbol{\gamma}}_i), \tag{A1}$$

where

$$\widehat{\mathbb{\Gamma}}_{in}(\hat{\boldsymbol{\gamma}}_i) = \frac{1}{n}\sum_{p=1}^{n}\left(\frac{\partial \boldsymbol{\psi}_i(Y_{pi}, \boldsymbol{X}_p; \hat{\boldsymbol{\gamma}}_i)}{\partial \boldsymbol{\gamma}_i^\top}\right),$$

$$\widehat{\mathbb{\Sigma}}_{in}(\boldsymbol{\gamma}_i) = \frac{1}{n}\sum_{p=1}^{n}\left(\boldsymbol{\psi}_i(Y_{pi}, \boldsymbol{X}_p; \hat{\boldsymbol{\gamma}}_i)\boldsymbol{\psi}_i^\top(Y_{pi}, \boldsymbol{X}_p; \hat{\boldsymbol{\gamma}}_i)\right),$$

with components of the matrix $\widehat{\mathbb{\Gamma}}_{in}(\boldsymbol{\gamma}_i) = \nabla^2 \text{RSS}_i(\boldsymbol{\gamma}_i)$ being

$$\begin{aligned}
\frac{\partial^2 \text{RSS}_i(\boldsymbol{\gamma}_i)}{\partial \gamma_{ik}\partial \gamma_{ij}} &= -2\sum_{p=1}^{n}\left\{(Y_{pi} - \pi_{pi})\frac{\partial^2 \pi_{pi}}{\partial \gamma_{ik}\partial \gamma_{ij}} - \frac{\partial \pi_{pi}}{\partial \gamma_{ik}}\frac{\partial \pi_{pi}}{\partial \gamma_{ij}}\right\},\\
\frac{\partial^2 \text{RSS}_i(\boldsymbol{\gamma}_i)}{\partial \gamma_{ik}^2} &= -2\sum_{p=1}^{n}\left\{(Y_{pi} - \pi_{pi})\frac{\partial^2 \pi_{pi}}{\partial \gamma_{ik}^2} - \left(\frac{\partial \pi_{pi}}{\partial \gamma_{ik}}\right)^2\right\}.
\end{aligned}$$

A.2 Maximum likelihood

Asymptotic properties of the ML estimator can be shown under the set of the following regularity conditions (van der Vaart, 1998, Theorems 5.41 and 5.42):

- [R0*] The support set $S = \{y \in \mathbb{R} : f(y|\boldsymbol{x}, \boldsymbol{\gamma}_i) > 0\}$ does not depend on the parameter $\boldsymbol{\gamma}_i$.
- [R1*] The true parameter $\boldsymbol{\gamma}_{iX}$ is an interior point of the parameter space.
- [R2*] The density $f(y|\boldsymbol{x}, \boldsymbol{\gamma}_i) = y\log(\pi(\boldsymbol{x}; \boldsymbol{\gamma}_i)) + (1-y)\log(1 - \pi(\boldsymbol{x}; \boldsymbol{\gamma}_i))$ is twice continuously differentiable with respect to $\boldsymbol{\gamma}_i$ for each $(y, \boldsymbol{x})$.
- [R3*] The Fisher information matrix $\mathbb{I}_i(\boldsymbol{\gamma}_i)$ is finite, regular, and positive definite in a neighborhood of γ_{iX}.
- [R4*] The order of differentiation and integration with respect to $\boldsymbol{\gamma}_i$ can be interchanged for terms $f(y|\boldsymbol{x}, \boldsymbol{\gamma}_i)$ and $\frac{\partial f(y|\boldsymbol{x}, \boldsymbol{\gamma}_i)}{\partial \boldsymbol{\gamma}_i}$.

Overall, the conditions [R0*] and [R2*] hold. In this case, the regularity condition for the ML estimator [R1*] is the same as condition [R1] for the NLS, so one must bound parameters of asymptotes to open intervals, as discussed in Section A.1.

For the condition [R3*], the Fisher information matrix takes the following form:

$$\mathbb{I}_i(\boldsymbol{\gamma}_i) = \mathsf{E}\,\mathbb{I}_i(\boldsymbol{\gamma}_i|\boldsymbol{X}_p) = \mathsf{E}\left(\frac{1}{\pi_{pi}(1-\pi_{pi})}\frac{\partial \pi_{pi}}{\partial \gamma_{ik}}\frac{\partial \pi_{pi}}{\partial \gamma_{ij}}\right)_{k,j},$$

which is a quadratic form and thus positive definite. Again, $\boldsymbol{X}_p$ describes respondent characteristics, so one can assume that their range is bounded, meaning partial derivatives $\frac{\partial \pi_{pi}}{\partial \gamma_{ik}}$, making the Fisher information matrix finite. Similarly, as in Section A.1, when the rows or columns of the Fisher information matrix $\mathbb{I}_i(\boldsymbol{\gamma}_i)$ are linearly independent, the matrix has a full rank and thus is regular, satisfying the condition [R3*]. The singularity of the matrix could occur in similar cases as for the matrix $\mathbb{\Gamma}_i(\boldsymbol{\gamma}_i)$ described in Section A.1.

Finally, regarding the condition [R4*], the order of differentiation and integration can be interchanged by dominated convergence theorem, as far as both $\frac{\partial f(y|\boldsymbol{x},\boldsymbol{\gamma}_i)}{\partial \boldsymbol{\gamma}_i}$ and $\frac{\partial^2 f(y|\boldsymbol{x},\boldsymbol{\gamma}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^\top}$ are dominated by an integrable function. In this case, a polynomial of x of the fourth degree can be taken as an integrable dominating function.

Hogg, McKean, and Craig (2018) demonstrated that when the regularity conditions [R0*]–[R4*] hold, there exists $n_0 \in \mathbf{N}$ and a sequence $\widehat{\boldsymbol{\gamma}}_{in}(n > n_0)$ of solutions to the corresponding likelihood equations such that

$$\widehat{\boldsymbol{\gamma}}_{in} \xrightarrow[n\to\infty]{P} \boldsymbol{\gamma}_{iX},$$

where $\boldsymbol{\gamma}_{iX}$ is a vector of true parameters. Because the log-likelihood function is not strictly concave, the approach described does not guarantee finding a unique solution to the corresponding likelihood equations. Thus, there might be multiple solutions, each a local maximum. However, there is one solution among them, which provides a consistent sequence of estimators. In contrast, other solutions may not even be close to γ_{iX} and may not converge to it. Therefore, in practice, the key part of estimating procedures is finding suitable starting values, preferably easily calculated but consistent estimates of parameters. Furthermore, for this consistent sequence of solutions, it can be shown that

$$\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{in} - \boldsymbol{\gamma}_{iX}\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbb{I}_i^{-1}(\boldsymbol{\gamma}_{iX})).$$

Estimate of asymptotic variance. An estimate of the asymptotic variance of the item parameters $\hat{\boldsymbol{\gamma}}_i$ is an inverse of the observed information matrix, an inverse of the Hessian matrix as demonstrated here:

$$\mathbb{I}_{in}^{-1}(\hat{\boldsymbol{\gamma}}_i|\boldsymbol{X},\boldsymbol{G}) = \left(-\frac{1}{n}\frac{\partial^2 l_i(\hat{\boldsymbol{\gamma}}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^\top}\right)^{-1}. \tag{A2}$$

B Tables

Table A1: Item parameters estimates with standard errors in parentheses by four estimation methods using the simple 4PL model (1) for the Anxiety dataset

Item	Method	b_0	b_1	d
R1	EM	$-0.76(-1.10, -0.42)$	$3.01(2.31, 3.71)$	$0.94(0.86, 1.01)$
	PLF	$-0.76(-1.10, -0.42)$	$3.00(2.30, 3.70)$	$0.94(0.86, 1.01)$
R2	EM	$-0.86(-1.26, -0.46)$	$2.89(2.07, 3.71)$	$0.90(0.80, 1.00)$
	PLF	$-0.87(-1.26, -0.47)$	$2.88(2.07, 3.70)$	$0.90(0.80, 1.00)$
R3	EM	$-1.55(-1.82, -1.28)$	$2.69(2.27, 3.11)$	$1.00(0.97, 1.03)$
	PLF	$-1.55(-1.82, -1.28)$	$2.69(2.27, 3.11)$	$1.00(0.97, 1.03)$
R4	EM	$1.17(0.25, 2.09)$	$3.86(2.40, 5.33)$	$0.98(0.91, 1.04)$
	PLF	$1.15(0.22, 2.08)$	$3.84(2.35, 5.33)$	$0.98(0.91, 1.05)$
R5	EM	$-1.63(-1.93, -1.34)$	$2.87(2.38, 3.35)$	$0.97(0.92, 1.02)$
	PLF	$-1.63(-1.93, -1.34)$	$2.87(2.38, 3.35)$	$0.97(0.92, 1.02)$
R6	EM	$-0.82(-1.04, -0.59)$	$2.42(2.05, 2.80)$	$1.00(0.97, 1.03)$
	PLF	$-0.82(-1.04, -0.59)$	$2.42(2.05, 2.80)$	$1.00(0.97, 1.03)$
R7	EM	$2.40(1.38, 3.41)$	$4.60(3.18, 6.02)$	$0.97(0.93, 1.01)$
	PLF	$2.39(1.38, 3.41)$	$4.60(3.17, 6.02)$	$0.97(0.93, 1.01)$
R8	EM	$-0.45(-0.80, -0.09)$	$2.05(1.55, 2.55)$	$0.88(0.79, 0.98)$
	PLF	$-0.45(-0.81, -0.10)$	$2.04(1.55, 2.54)$	$0.88(0.79, 0.98)$
R9	EM	$0.41(-0.14, 0.97)$	$2.52(1.75, 3.29)$	$0.84(0.76, 0.93)$

Table A1: continued from previous page

Item	Method	b_0	b_1	d
	PLF	$0.41(-0.15, 0.96)$	$2.51(1.75, 3.28)$	$0.84(0.76, 0.93)$
R10	EM	$-1.60(-1.90, -1.30)$	$3.06(2.50, 3.63)$	$0.98(0.93, 1.03)$
	PLF	$-1.60(-1.90, -1.30)$	$3.06(2.50, 3.63)$	$0.98(0.93, 1.03)$
R11	EM	$0.27(-0.20, 0.73)$	$2.47(1.79, 3.16)$	$0.90(0.82, 0.98)$
	PLF	$0.26(-0.21, 0.73)$	$2.47(1.79, 3.15)$	$0.90(0.82, 0.98)$
R12	EM	$2.00(1.24, 2.75)$	$4.26(3.18, 5.35)$	$0.93(0.89, 0.98)$
	PLF	$2.00(1.25, 2.75)$	$4.27(3.18, 5.35)$	$0.93(0.89, 0.98)$
R13	EM	$-0.51(-0.92, -0.09)$	$2.03(1.40, 2.65)$	$0.99(0.86, 1.11)$
	PLF	$-0.50(-0.92, -0.08)$	$2.04(1.41, 2.67)$	$0.98(0.86, 1.11)$
R14	EM	$0.34(0.09, 0.59)$	$2.62(2.20, 3.05)$	$1.00(0.98, 1.02)$
	PLF	$0.34(0.09, 0.59)$	$2.62(2.20, 3.05)$	$1.00(0.98, 1.02)$
R15	EM	$-0.98(-1.22, -0.74)$	$2.24(1.87, 2.62)$	$1.00(0.95, 1.05)$
	PLF	$-0.98(-1.22, -0.74)$	$2.24(1.87, 2.62)$	$1.00(0.95, 1.05)$
R16	EM	$2.49(1.90, 3.08)$	$4.90(4.01, 5.80)$	$0.98(0.97, 1.00)$
	PLF	$2.49(1.90, 3.08)$	$4.91(4.01, 5.80)$	$0.98(0.97, 1.00)$
R17	EM	$-2.75(-3.18, -2.32)$	$2.54(1.99, 3.09)$	$0.93(0.82, 1.04)$
	PLF	$-2.75(-3.18, -2.32)$	$2.54(1.99, 3.09)$	$0.93(0.82, 1.04)$
R18	EM	$1.51(0.83, 2.19)$	$3.37(2.45, 4.29)$	$0.90(0.84, 0.95)$
	PLF	$1.51(0.83, 2.19)$	$3.36(2.45, 4.28)$	$0.90(0.84, 0.95)$
R19	EM	$-1.55(-1.83, -1.27)$	$2.80(2.32, 3.27)$	$0.99(0.95, 1.03)$
	PLF	$-1.55(-1.83, -1.27)$	$2.80(2.32, 3.28)$	$0.99(0.95, 1.03)$
R20	EM	$-0.93(-1.20, -0.67)$	$2.83(2.29, 3.37)$	$0.99(0.95, 1.04)$
	PLF	$-0.93(-1.20, -0.67)$	$2.83(2.29, 3.38)$	$0.99(0.95, 1.04)$
R21	EM	$-0.06(-0.65, 0.53)$	$1.95(1.23, 2.68)$	$0.76(0.63, 0.89)$
	PLF	$-0.07(-0.65, 0.52)$	$1.95(1.23, 2.66)$	$0.76(0.63, 0.89)$
R22	EM	$1.16(0.61, 1.71)$	$4.12(3.17, 5.08)$	$0.98(0.95, 1.02)$
	PLF	$1.16(0.62, 1.71)$	$4.13(3.17, 5.08)$	$0.98(0.95, 1.02)$
R23	EM	$1.06(0.48, 1.64)$	$3.01(2.17, 3.85)$	$0.96(0.91, 1.01)$
	PLF	$1.06(0.48, 1.64)$	$3.01(2.17, 3.85)$	$0.96(0.91, 1.01)$
R24	EM	$1.52(0.87, 2.16)$	$4.38(3.34, 5.42)$	$0.93(0.89, 0.98)$
	PLF	$1.51(0.87, 2.16)$	$4.38(3.34, 5.41)$	$0.93(0.89, 0.98)$
R25	EM	$3.30(2.14, 4.46)$	$4.40(2.99, 5.81)$	$0.93(0.90, 0.97)$
	PLF	$3.29(2.13, 4.45)$	$4.40(2.99, 5.81)$	$0.93(0.90, 0.97)$
R26	EM	$2.55(1.68, 3.41)$	$4.96(3.73, 6.20)$	$0.95(0.92, 0.99)$
	PLF	$2.55(1.68, 3.42)$	$4.97(3.73, 6.20)$	$0.95(0.92, 0.99)$
R27	EM	$0.93(0.63, 1.23)$	$3.64(3.09, 4.19)$	$1.00(0.99, 1.01)$
	PLF	$0.93(0.63, 1.23)$	$3.64(3.09, 4.19)$	$1.00(0.99, 1.01)$
R28	EM	$2.60(1.87, 3.32)$	$5.06(3.99, 6.14)$	$0.98(0.95, 1.00)$
	PLF	$2.59(1.87, 3.32)$	$5.06(3.99, 6.13)$	$0.98(0.95, 1.00)$
R29	EM	$-0.13(-0.62, 0.36)$	$4.08(2.95, 5.22)$	$0.91(0.85, 0.98)$
	PLF	$-0.13(-0.61, 0.36)$	$4.09(2.95, 5.23)$	$0.91(0.85, 0.98)$

Table A2: Item parameters estimates with confidence intervals in parentheses by two estimation methods using the group-specific 4PL model (3) for the Anxiety dataset

Item	Method	b_0	b_1	b_2	b_3	d	d_{DIF}
R1	EM	$-0.73(-1.28, -0.18)$	$3.39(2.25, 4.54)$	$-0.10(-0.83, 0.62)$	$-0.87(-2.38, 0.64)$	$0.91(0.80, 1.02)$	$0.07(-0.10, 0.25)$
	PLF	$-0.73(-1.28, -0.18)$	$3.39(2.25, 4.52)$	$-0.09(-0.81, 0.63)$	$-0.85(-2.35, 0.65)$	$0.92(0.81, 1.02)$	$0.07(-0.10, 0.24)$
R2	EM	$-0.70(-1.53, 0.14)$	$3.52(1.74, 5.30)$	$-0.29(-1.18, 0.61)$	$-1.16(-3.03, 0.70)$	$0.81(0.65, 0.97)$	$0.19(0.01, 0.36)$
	PLF	$-0.69(-1.53, 0.14)$	$3.52(1.74, 5.30)$	$-0.29(-1.19, 0.60)$	$-1.17(-3.03, 0.69)$	$0.81(0.65, 0.97)$	$0.19(0.02, 0.36)$
R3	EM	$-1.42(-1.88, -0.95)$	$2.64(1.76, 3.52)$	$-0.23(-0.84, 0.38)$	$0.26(-0.81, 1.32)$	$0.97(0.84, 1.10)$	$0.03(-0.11, 0.16)$
	PLF	$-1.42(-1.88, -0.95)$	$2.65(1.76, 3.53)$	$-0.23(-0.84, 0.38)$	$0.25(-0.81, 1.32)$	$0.97(0.84, 1.10)$	$0.03(-0.11, 0.16)$
R4	EM	$0.77(0.34, 1.20)$	$3.09(2.37, 3.80)$	$1.02(-0.09, 2.13)$	$2.12(0.16, 4.08)$	$1.00(0.98, 1.02)$	$-0.05(-0.10, 0.01)$
	PLF	$0.77(0.35, 1.20)$	$3.08(2.37, 3.80)$	$1.02(-0.09, 2.12)$	$2.11(0.15, 4.07)$	$1.00(0.98, 1.02)$	$-0.05(-0.10, 0.01)$
R5	EM	$-1.67(-2.07, -1.26)$	$2.61(2.02, 3.20)$	$0.01(-0.57, 0.60)$	$0.47(-0.47, 1.41)$	$1.00(0.95, 1.05)$	$-0.05(-0.14, 0.04)$
	PLF	$-1.67(-2.07, -1.26)$	$2.61(2.02, 3.20)$	$0.01(-0.57, 0.60)$	$0.47(-0.47, 1.41)$	$1.00(0.95, 1.05)$	$-0.05(-0.14, 0.04)$
R6	EM	$-0.46(-0.78, -0.13)$	$2.26(1.76, 2.75)$	$-0.79(-1.28, -0.30)$	$0.91(-0.03, 1.84)$	$1.00(0.95, 1.05)$	$-0.03(-0.10, 0.05)$
	PLF	$-0.46(-0.78, -0.13)$	$2.26(1.76, 2.75)$	$-0.79(-1.28, -0.30)$	$0.91(-0.02, 1.84)$	$1.00(0.95, 1.05)$	$-0.03(-0.10, 0.05)$
R7	EM	$3.99(2.18, 5.80)$	$6.79(4.30, 9.29)$	$-2.22(-4.10, -0.34)$	$-3.08(-5.70, -0.45)$	$0.92(0.87, 0.98)$	$0.08(0.02, 0.13)$
	PLF	$4.00(2.19, 5.81)$	$6.80(4.31, 9.30)$	$-2.22(-4.11, -0.34)$	$-3.08(-5.71, -0.45)$	$0.92(0.87, 0.98)$	$0.08(0.02, 0.13)$
R8	EM	$-0.40(-0.96, 0.16)$	$2.11(1.35, 2.88)$	$-0.09(-0.81, 0.64)$	$-0.13(-1.14, 0.89)$	$0.86(0.72, 1.01)$	$0.03(-0.16, 0.22)$
	PLF	$-0.40(-0.96, 0.15)$	$2.11(1.35, 2.87)$	$-0.08(-0.81, 0.64)$	$-0.12(-1.14, 0.89)$	$0.87(0.72, 1.01)$	$0.03(-0.16, 0.22)$
R9	EM	$0.50(-0.26, 1.26)$	$2.45(1.43, 3.47)$	$-0.16(-1.26, 0.93)$	$0.28(-1.27, 1.84)$	$0.89(0.77, 1.01)$	$-0.08(-0.25, 0.09)$
	PLF	$0.50(-0.26, 1.26)$	$2.44(1.43, 3.46)$	$-0.17(-1.26, 0.93)$	$0.28(-1.27, 1.83)$	$0.89(0.77, 1.01)$	$-0.08(-0.25, 0.09)$
R10	EM	$-0.98(-1.72, -0.24)$	$3.63(1.76, 5.49)$	$-1.05(-1.92, -0.18)$	$-0.37(-2.38, 1.64)$	$0.89(0.73, 1.06)$	$0.11(-0.06, 0.27)$
	PLF	$-0.99(-1.72, -0.26)$	$3.61(1.77, 5.45)$	$-1.05(-1.91, -0.18)$	$-0.35(-2.34, 1.63)$	$0.89(0.73, 1.06)$	$0.11(-0.06, 0.27)$
R11	EM	$0.49(-0.38, 1.35)$	$2.72(1.49, 3.95)$	$-0.30(-1.28, 0.69)$	$-0.30(-1.74, 1.13)$	$0.84(0.72, 0.97)$	$0.09(-0.06, 0.24)$
	PLF	$0.48(-0.39, 1.34)$	$2.71(1.48, 3.93)$	$-0.29(-1.28, 0.70)$	$-0.30(-1.72, 1.13)$	$0.84(0.72, 0.97)$	$0.09(-0.06, 0.24)$
R12	EM	$1.88(1.12, 2.64)$	$4.03(2.91, 5.15)$	$0.22(-1.17, 1.61)$	$0.55(-1.52, 2.61)$	$0.98(0.94, 1.02)$	$-0.07(-0.14, -0.00)$
	PLF	$1.88(1.12, 2.64)$	$4.03(2.91, 5.15)$	$0.22(-1.17, 1.61)$	$0.55(-1.52, 2.61)$	$0.98(0.94, 1.02)$	$-0.07(-0.14, -0.00)$
R13	EM	$-0.39(-0.75, -0.02)$	$2.16(1.59, 2.72)$	$-0.22(-0.91, 0.47)$	$-0.19(-1.27, 0.88)$	$1.00(0.94, 1.06)$	$-0.02(-0.23, 0.18)$
	PLF	$-0.39(-0.74, -0.03)$	$2.16(1.60, 2.71)$	$-0.23(-0.92, 0.46)$	$-0.21(-1.28, 0.85)$	$1.00(0.94, 1.06)$	$-0.02(-0.23, 0.19)$
R14	EM	$0.26(-0.11, 0.63)$	$2.54(1.93, 3.15)$	$0.14(-0.36, 0.65)$	$0.15(-0.70, 1.00)$	$1.00(0.97, 1.03)$	$0.00(-0.04, 0.04)$
	PLF	$0.26(-0.11, 0.63)$	$2.54(1.93, 3.15)$	$0.14(-0.36, 0.65)$	$0.15(-0.70, 1.00)$	$1.00(0.97, 1.03)$	$0.00(-0.04, 0.04)$
R15	EM	$-1.08(-1.45, -0.72)$	$2.09(1.55, 2.62)$	$0.18(-0.30, 0.67)$	$0.28(-0.46, 1.01)$	$1.00(0.91, 1.09)$	$0.00(-0.10, 0.10)$
	PLF	$-1.08(-1.45, -0.72)$	$2.09(1.55, 2.62)$	$0.18(-0.30, 0.67)$	$0.28(-0.46, 1.01)$	$1.00(0.91, 1.09)$	$0.00(-0.10, 0.10)$
R16	EM	$2.37(1.56, 3.18)$	$4.92(3.68, 6.16)$	$0.24(-0.94, 1.42)$	$-0.03(-1.82, 1.75)$	$0.99(0.96, 1.01)$	$-0.00(-0.04, 0.04)$
	PLF	$2.37(1.56, 3.18)$	$4.92(3.68, 6.16)$	$0.24(-0.94, 1.42)$	$-0.04(-1.82, 1.75)$	$0.99(0.96, 1.01)$	$-0.00(-0.04, 0.04)$
R17	EM	$-3.17(-3.98, -2.36)$	$3.14(2.00, 4.29)$	$0.64(-0.31, 1.58)$	$-1.05(-2.33, 0.23)$	$0.89(0.71, 1.06)$	$0.11(-0.11, 0.33)$
	PLF	$-3.17(-3.98, -2.36)$	$3.14(2.00, 4.29)$	$0.64(-0.31, 1.58)$	$-1.05(-2.33, 0.22)$	$0.89(0.71, 1.06)$	$0.11(-0.11, 0.33)$
R18	EM	$1.57(0.20, 2.94)$	$3.31(1.57, 5.04)$	$-0.04(-1.65, 1.56)$	$0.24(-1.86, 2.33)$	$0.88(0.77, 1.00)$	$0.02(-0.11, 0.15)$
	PLF	$1.56(0.22, 2.90)$	$3.29(1.59, 4.98)$	$-0.03(-1.61, 1.55)$	$0.25(-1.82, 2.32)$	$0.88(0.77, 1.00)$	$0.02(-0.11, 0.15)$
R19	EM	$-1.39(-1.77, -1.00)$	$2.77(2.15, 3.39)$	$-0.33(-0.89, 0.23)$	$0.14(-0.79, 1.07)$	$1.00(0.96, 1.04)$	$-0.02(-0.08, 0.05)$
	PLF	$-1.39(-1.77, -1.00)$	$2.77(2.15, 3.39)$	$-0.33(-0.89, 0.23)$	$0.14(-0.79, 1.07)$	$1.00(0.96, 1.04)$	$-0.02(-0.08, 0.05)$
R20	EM	$-1.09(-1.45, -0.73)$	$2.76(2.13, 3.39)$	$0.29(-0.21, 0.79)$	$0.12(-0.85, 1.08)$	$1.00(0.96, 1.04)$	$-0.01(-0.08, 0.06)$
	PLF	$-1.09(-1.45, -0.73)$	$2.76(2.13, 3.39)$	$0.29(-0.21, 0.79)$	$0.11(-0.85, 1.08)$	$1.00(0.96, 1.04)$	$-0.01(-0.08, 0.06)$
R21	EM	$0.03(-0.80, 0.85)$	$1.95(0.95, 2.95)$	$-0.19(-1.34, 0.96)$	$0.13(-1.32, 1.58)$	$0.84(0.66, 1.02)$	$-0.14(-0.39, 0.10)$
	PLF	$0.02(-0.80, 0.84)$	$1.95(0.95, 2.94)$	$-0.19(-1.34, 0.95)$	$0.12(-1.32, 1.57)$	$0.84(0.66, 1.02)$	$-0.14(-0.39, 0.11)$
R22	EM	$0.93(0.45, 1.41)$	$4.02(3.12, 4.92)$	$0.06(-0.60, 0.72)$	$-0.47(-1.69, 0.74)$	$1.00(0.98, 1.02)$	$0.00(-0.02, 0.02)$
	PLF	$0.94(0.43, 1.45)$	$4.04(3.08, 4.99)$	$0.31(-0.57, 1.19)$	$-0.05(-1.62, 1.51)$	$1.00(0.98, 1.02)$	$-0.02(-0.07, 0.03)$
R23	EM	$1.34(0.54, 2.15)$	$3.63(2.44, 4.81)$	$-0.61(-1.49, 0.27)$	$-1.27(-2.57, 0.04)$	$0.94(0.87, 1.01)$	$0.06(-0.01, 0.14)$
	PLF	$1.34(0.54, 2.15)$	$3.62(2.44, 4.81)$	$-0.61(-1.49, 0.27)$	$-1.27(-2.57, 0.04)$	$0.94(0.87, 1.01)$	$0.06(-0.01, 0.14)$
R24	EM	$1.21(0.55, 1.87)$	$4.11(2.98, 5.25)$	$0.87(-0.49, 2.23)$	$0.94(-1.25, 3.14)$	$0.98(0.94, 1.02)$	$-0.09(-0.16, -0.02)$
	PLF	$1.21(0.55, 1.87)$	$4.11(2.98, 5.25)$	$0.87(-0.49, 2.23)$	$0.94(-1.25, 3.14)$	$0.98(0.94, 1.02)$	$-0.09(-0.16, -0.02)$
R25	EM	$3.71(0.62, 6.80)$	$4.83(1.19, 8.47)$	$-0.45(-3.80, 2.89)$	$-0.42(-4.39, 3.55)$	$0.92(0.84, 1.00)$	$0.02(-0.07, 0.11)$
	PLF	$3.67(0.60, 6.75)$	$4.79(1.17, 8.41)$	$-0.42(-3.75, 2.91)$	$-0.38(-4.33, 3.58)$	$0.92(0.84, 1.00)$	$0.02(-0.07, 0.11)$
R26	EM	$2.87(1.27, 4.47)$	$5.43(3.20, 7.66)$	$-0.74(-2.62, 1.14)$	$-1.04(-3.70, 1.62)$	$0.97(0.91, 1.03)$	$-0.02(-0.10, 0.05)$
	PLF	$2.86(1.28, 4.44)$	$5.41(3.20, 7.63)$	$-0.73(-2.60, 1.13)$	$-1.02(-3.67, 1.62)$	$0.97(0.91, 1.03)$	$-0.02(-0.10, 0.05)$
R27	EM	$1.08(0.60, 1.57)$	$3.91(3.04, 4.79)$	$-0.26(-0.89, 0.36)$	$-0.47(-1.59, 0.66)$	$1.00(0.98, 1.02)$	$0.00(-0.02, 0.02)$
	PLF	$1.08(0.60, 1.57)$	$3.91(3.04, 4.79)$	$-0.26(-0.89, 0.36)$	$-0.47(-1.59, 0.66)$	$1.00(0.98, 1.02)$	$0.00(-0.02, 0.02)$
R28	EM	$2.12(1.26, 2.98)$	$4.54(3.25, 5.82)$	$0.92(-0.44, 2.28)$	$1.00(-1.02, 3.02)$	$0.98(0.95, 1.02)$	$-0.01(-0.06, 0.04)$
	PLF	$2.12(1.26, 2.98)$	$4.54(3.25, 5.83)$	$0.92(-0.44, 2.28)$	$1.00(-1.02, 3.02)$	$0.98(0.95, 1.02)$	$-0.01(-0.06, 0.04)$
R29	EM	$0.39(-0.21, 0.99)$	$4.90(3.52, 6.28)$	$-1.07(-1.82, -0.32)$	$-1.69(-3.44, 0.06)$	$0.91(0.84, 0.98)$	$0.06(-0.05, 0.18)$

Table A2: continued from previous page

Item	Method	b_0	b_1	b_2	b_3	d	d_{DIF}
	PLF	$0.39(-0.21, 0.99)$	$4.90(3.52, 6.28)$	$-1.07(-1.82, -0.32)$	$-1.68(-3.43, 0.08)$	$0.91(0.84, 0.98)$	$0.06(-0.05, 0.17)$

Table A3: Item parameters estimates with standard errors in parentheses by four estimation methods using the simple 4PL model (1) for the LearningToLearn dataset

Item	Method	b_0	b_1	c
1A	EM	$2.01(0.76, 3.26)$	$1.00(0.54, 1.46)$	$0.01(-1.05, 1.08)$
	PLF	$1.92(0.88, 2.95)$	$1.03(0.61, 1.44)$	$0.09(-0.69, 0.88)$
1B	EM	$1.64(0.26, 3.01)$	$0.82(0.48, 1.17)$	$0.00(-1.09, 1.09)$
	PLF	$1.64(0.26, 3.01)$	$0.82(0.48, 1.17)$	$0.00(-1.09, 1.09)$
1C	EM	$1.45(0.54, 2.36)$	$0.74(0.51, 0.96)$	$0.00(-0.69, 0.69)$
	PLF	$1.45(0.54, 2.36)$	$0.74(0.51, 0.96)$	$0.00(-0.69, 0.69)$
1D	EM	$0.17(-0.43, 0.77)$	$0.87(0.55, 1.18)$	$0.00(-0.28, 0.28)$
	PLF	$0.17(-0.43, 0.77)$	$0.87(0.55, 1.18)$	$0.00(-0.28, 0.28)$
1E	EM	$-0.16(-1.06, 0.74)$	$0.84(0.41, 1.26)$	$0.17(-0.12, 0.47)$
	PLF	$-0.15(-1.05, 0.75)$	$0.83(0.41, 1.26)$	$0.17(-0.13, 0.47)$
1F	EM	$-0.02(-0.55, 0.51)$	$0.87(0.61, 1.13)$	$0.00(-0.23, 0.23)$
	PLF	$-0.02(-0.55, 0.51)$	$0.87(0.61, 1.13)$	$0.00(-0.23, 0.23)$
1G	EM	$-0.37(-0.91, 0.16)$	$0.87(0.58, 1.16)$	$0.00(-0.18, 0.18)$
	PLF	$-0.37(-0.91, 0.16)$	$0.87(0.58, 1.16)$	$0.00(-0.18, 0.18)$
1H	EM	$-1.25(-2.18, -0.32)$	$0.48(0.20, 0.75)$	$0.00(-0.20, 0.20)$
	PLF	$-1.25(-2.18, -0.32)$	$0.48(0.20, 0.75)$	$0.00(-0.20, 0.20)$
2A	EM	$3.94(1.94, 5.93)$	$1.11(0.66, 1.57)$	$0.00(-1.82, 1.82)$
	PLF	$3.94(1.94, 5.93)$	$1.11(0.66, 1.57)$	$0.00(-1.82, 1.82)$
2B	EM	$3.28(1.46, 5.11)$	$0.90(0.54, 1.27)$	$0.00(-1.67, 1.67)$
	PLF	$3.28(1.46, 5.11)$	$0.90(0.54, 1.27)$	$0.00(-1.67, 1.67)$
2C	EM	$2.91(1.38, 4.45)$	$1.07(0.69, 1.45)$	$0.00(-1.40, 1.40)$
	PLF	$2.91(1.38, 4.45)$	$1.07(0.69, 1.45)$	$0.00(-1.40, 1.40)$
2D	EM	$2.24(1.32, 3.17)$	$1.06(0.77, 1.34)$	$0.00(-0.76, 0.76)$
	PLF	$2.24(1.32, 3.17)$	$1.06(0.77, 1.34)$	$0.00(-0.76, 0.76)$
2E	EM	$1.07(0.05, 2.08)$	$0.73(0.46, 0.99)$	$0.00(-0.70, 0.70)$
	PLF	$1.07(0.05, 2.08)$	$0.73(0.46, 0.99)$	$0.00(-0.70, 0.70)$
2F	EM	$1.28(-0.12, 2.68)$	$0.56(0.30, 0.82)$	$0.00(-1.06, 1.06)$
	PLF	$1.28(-0.12, 2.68)$	$0.56(0.30, 0.82)$	$0.00(-1.06, 1.06)$
2G	EM	$-0.89(-1.65, -0.12)$	$0.62(0.28, 0.96)$	$0.00(-0.20, 0.20)$
	PLF	$-0.89(-1.66, -0.12)$	$0.62(0.28, 0.96)$	$0.00(-0.19, 0.20)$
3	EM	$0.61(-0.69, 1.90)$	$0.57(0.27, 0.87)$	$0.00(-0.80, 0.80)$
	PLF	$0.61(-0.69, 1.90)$	$0.57(0.27, 0.87)$	$0.00(-0.80, 0.80)$
4A	EM	$-0.87(-1.39, -0.35)$	$1.10(0.77, 1.43)$	$0.00(-0.12, 0.12)$
	PLF	$-0.87(-1.39, -0.35)$	$1.10(0.77, 1.43)$	$0.00(-0.12, 0.12)$
4B	EM	$-0.53(-1.02, -0.03)$	$1.22(0.89, 1.56)$	$0.00(-0.14, 0.14)$
	PLF	$-0.53(-1.02, -0.03)$	$1.22(0.89, 1.56)$	$0.00(-0.14, 0.14)$
4C	EM	$-1.20(-1.61, -0.78)$	$1.39(1.03, 1.76)$	$0.00(-0.05, 0.06)$
	PLF	$-1.20(-1.62, -0.79)$	$1.40(1.03, 1.76)$	$0.01(-0.05, 0.06)$
4D	EM	$-1.08(-1.34, -0.82)$	$1.28(1.02, 1.53)$	$0.01(-0.02, 0.04)$
	PLF	$-1.08(-1.34, -0.82)$	$1.28(1.02, 1.53)$	$0.01(-0.02, 0.04)$

Table A3: continued from previous page

Item	Method	b_0	b_1	d
5A	EM	$1.43(-0.13, 3.00)$	$0.70(0.36, 1.04)$	$0.00(-1.21, 1.21)$
	PLF	$1.43(-0.13, 3.00)$	$0.70(0.36, 1.04)$	$0.00(-1.21, 1.21)$
5B	EM	$0.82(-1.05, 2.69)$	$0.41(0.13, 0.69)$	$0.00(-1.27, 1.27)$
	PLF	$0.82(-1.05, 2.69)$	$0.41(0.13, 0.69)$	$0.00(-1.27, 1.27)$
5C	EM	$-0.55(-2.84, 1.74)$	$0.40(-0.14, 0.95)$	$0.27(-0.32, 0.85)$
	PLF	$-0.46(-3.29, 2.37)$	$0.38(-0.23, 1.00)$	$0.24(-0.56, 1.04)$
5D	EM	$0.27(-1.00, 1.54)$	$0.92(0.34, 1.50)$	$0.35(-0.06, 0.76)$
	PLF	$0.25(-1.04, 1.53)$	$0.93(0.34, 1.52)$	$0.36(-0.05, 0.76)$
5E	EM	$1.52(-1.93, 4.98)$	$0.62(0.07, 1.17)$	$0.00(-2.76, 2.76)$
	PLF	$1.52(-2.68, 5.73)$	$0.62(-0.03, 1.27)$	$0.00(-3.36, 3.36)$
5F	EM	$-1.81(-8.92, 5.30)$	$0.31(-1.39, 2.02)$	$0.18(-0.61, 0.97)$
	PLF	$-1.82(-8.84, 5.20)$	$0.32(-1.38, 2.01)$	$0.18(-0.59, 0.95)$
5G	EM	$-5.96(-12.03, 0.11)$	$1.81(-0.67, 4.29)$	$0.11(0.08, 0.15)$
	PLF	$-5.94(-11.98, 0.10)$	$1.80(-0.67, 4.27)$	$0.11(0.08, 0.15)$
6A	EM	$-1.22(-1.66, -0.78)$	$1.39(1.02, 1.75)$	$0.03(-0.04, 0.09)$
	PLF	$-1.23(-1.67, -0.78)$	$1.39(1.03, 1.76)$	$0.03(-0.04, 0.09)$
6B	EM	$-1.47(-2.25, -0.69)$	$1.13(0.67, 1.59)$	$0.00(-0.11, 0.11)$
	PLF	$-1.47(-2.25, -0.69)$	$1.13(0.67, 1.59)$	$0.00(-0.11, 0.11)$
6C	EM	$-4.21(-6.20, -2.22)$	$1.60(0.61, 2.59)$	$0.04(0.01, 0.07)$
	PLF	$-4.23(-6.23, -2.22)$	$1.61(0.61, 2.60)$	$0.04(0.01, 0.07)$
6D	EM	$-0.60(-1.20, -0.00)$	$1.23(0.79, 1.68)$	$0.04(-0.11, 0.19)$
	PLF	$-0.62(-1.22, -0.02)$	$1.24(0.79, 1.69)$	$0.04(-0.10, 0.19)$
6E	EM	$-1.98(-2.68, -1.28)$	$1.23(0.79, 1.68)$	$0.07(0.01, 0.13)$
	PLF	$-1.99(-2.70, -1.28)$	$1.24(0.79, 1.68)$	$0.07(0.01, 0.13)$
6F	EM	$-1.63(-2.51, -0.75)$	$1.19(0.65, 1.74)$	$0.07(-0.04, 0.17)$
	PLF	$-1.61(-2.47, -0.74)$	$1.18(0.64, 1.71)$	$0.06(-0.04, 0.17)$
6G	EM	$-2.35(-3.52, -1.18)$	$1.03(0.41, 1.65)$	$0.09(0.01, 0.16)$
	PLF	$-2.50(-3.71, -1.29)$	$1.10(0.47, 1.74)$	$0.10(0.02, 0.17)$
6H	EM	$-1.69(-3.19, -0.20)$	$0.78(0.09, 1.48)$	$0.06(-0.12, 0.25)$
	PLF	$-1.67(-3.16, -0.18)$	$0.77(0.08, 1.46)$	$0.06(-0.13, 0.25)$
7A	EM	$-0.14(-0.88, 0.60)$	$0.92(0.53, 1.31)$	$0.00(-0.29, 0.29)$
	PLF	$-0.14(-0.88, 0.60)$	$0.92(0.53, 1.31)$	$0.00(-0.29, 0.29)$
7B	EM	$-0.10(-1.27, 1.08)$	$0.83(0.31, 1.35)$	$0.30(-0.04, 0.64)$
	PLF	$-0.08(-1.27, 1.11)$	$0.82(0.30, 1.34)$	$0.29(-0.06, 0.65)$
7C	EM	$-0.23(-0.88, 0.43)$	$0.78(0.46, 1.10)$	$0.00(-0.25, 0.25)$
	PLF	$-0.23(-0.88, 0.43)$	$0.78(0.46, 1.10)$	$0.00(-0.25, 0.25)$
7D	EM	$0.29(-0.33, 0.91)$	$0.83(0.53, 1.14)$	$0.00(-0.31, 0.31)$
	PLF	$0.29(-0.33, 0.91)$	$0.83(0.53, 1.14)$	$0.00(-0.31, 0.31)$
7E	EM	$-1.56(-2.89, -0.24)$	$0.73(0.12, 1.35)$	$0.03(-0.16, 0.22)$
	PLF	$-1.57(-2.89, -0.25)$	$0.74(0.12, 1.35)$	$0.03(-0.16, 0.22)$
7F	EM	$-1.09(-2.22, 0.04)$	$0.50(0.14, 0.87)$	$0.00(-0.26, 0.26)$
	PLF	$-1.09(-2.22, 0.04)$	$0.50(0.14, 0.87)$	$0.00(-0.26, 0.26)$

Table A4: Item parameters estimates with confidence intervals in parentheses by two estimation methods using the group-specific 4PL model (3) for the LearningToLearn dataset

Item	Method	b_0	b_1	b_2	b_3	c	c_{DIF}
1A	EM	$2.02(1.15, 2.88)$	$1.36(0.85, 1.87)$	$-0.94(-3.50, 1.63)$	$-0.50(-1.45, 0.44)$	$0.00(-0.67, 0.68)$	$0.55(-0.45, 1.56)$
	PLF	$2.02(1.15, 2.88)$	$1.36(0.85, 1.87)$	$-0.92(-3.50, 1.65)$	$-0.51(-1.45, 0.44)$	$0.00(-0.68, 0.68)$	$0.55(-0.47, 1.57)$
1B	EM	$1.62(0.30, 2.94)$	$0.85(0.48, 1.22)$	$-0.31(-3.03, 2.42)$	$0.03(-0.81, 0.88)$	$0.00(-1.02, 1.02)$	$0.23(-1.46, 1.93)$
	PLF	$1.62(0.15, 3.09)$	$0.85(0.46, 1.24)$	$-0.43(-3.11, 2.25)$	$0.07(-0.80, 0.93)$	$0.00(-1.14, 1.14)$	$0.30(-1.30, 1.90)$
1C	EM	$1.39(0.07, 2.70)$	$0.78(0.43, 1.13)$	$0.14(-1.71, 1.98)$	$-0.08(-0.55, 0.38)$	$0.00(-0.98, 0.98)$	$-0.00(-1.39, 1.39)$
	PLF	$1.39(0.07, 2.70)$	$0.78(0.43, 1.13)$	$0.14(-1.71, 1.98)$	$-0.08(-0.55, 0.38)$	$0.00(-0.98, 0.98)$	$0.00(-1.39, 1.39)$
1D	EM	$-0.09(-0.74, 0.56)$	$0.87(0.48, 1.26)$	$0.52(-0.84, 1.88)$	$0.02(-0.67, 0.71)$	$0.00(-0.25, 0.25)$	$0.01(-0.67, 0.68)$
	PLF	$-0.09(-0.74, 0.56)$	$0.87(0.48, 1.26)$	$0.51(-0.85, 1.88)$	$0.02(-0.67, 0.72)$	$0.00(-0.25, 0.25)$	$0.01(-0.67, 0.69)$
1E	EM	$-0.26(-1.65, 1.12)$	$0.83(0.17, 1.49)$	$0.23(-1.61, 2.07)$	$0.01(-0.86, 0.88)$	$0.13(-0.32, 0.58)$	$0.08(-0.53, 0.69)$
	PLF	$-0.29(-1.68, 1.09)$	$0.85(0.18, 1.51)$	$0.21(-1.64, 2.06)$	$0.01(-0.87, 0.90)$	$0.14(-0.29, 0.57)$	$0.09(-0.50, 0.67)$
1F	EM	$-0.14(-0.84, 0.55)$	$1.03(0.64, 1.42)$	$0.23(-0.90, 1.37)$	$-0.30(-0.83, 0.22)$	$0.00(-0.27, 0.27)$	$-0.00(-0.50, 0.50)$
	PLF	$-0.14(-0.84, 0.55)$	$1.03(0.64, 1.42)$	$0.23(-0.90, 1.37)$	$-0.30(-0.83, 0.22)$	$0.00(-0.27, 0.27)$	$0.00(-0.50, 0.50)$
1G	EM	$-0.46(-1.20, 0.28)$	$0.93(0.52, 1.35)$	$0.17(-0.91, 1.25)$	$-0.12(-0.70, 0.46)$	$0.00(-0.23, 0.23)$	$-0.00(-0.37, 0.37)$
	PLF	$-0.46(-1.20, 0.28)$	$0.93(0.52, 1.35)$	$0.17(-0.91, 1.25)$	$-0.12(-0.70, 0.46)$	$0.00(-0.23, 0.23)$	$0.00(-0.37, 0.37)$
1H	EM	$-1.39(-3.12, 0.34)$	$0.40(-0.05, 0.86)$	$0.25(-1.77, 2.28)$	$0.15(-0.43, 0.72)$	$0.00(-0.33, 0.33)$	$-0.00(-0.41, 0.41)$
	PLF	$-1.37(-3.10, 0.35)$	$0.40(-0.05, 0.85)$	$0.24(-1.79, 2.26)$	$0.15(-0.42, 0.72)$	$0.00(-0.34, 0.34)$	$0.00(-0.41, 0.41)$
2A	EM	$4.15(-0.24, 8.54)$	$1.08(0.35, 1.81)$	$-0.38(-5.30, 4.53)$	$0.06(-0.88, 1.01)$	$0.00(-4.21, 4.21)$	$-0.00(-4.63, 4.63)$
	PLF	$4.15(-0.24, 8.54)$	$1.08(0.35, 1.81)$	$-0.38(-5.30, 4.53)$	$0.06(-0.88, 1.01)$	$0.00(-4.21, 4.21)$	$0.00(-4.63, 4.63)$
2B	EM	$3.22(-1.88, 8.32)$	$0.67(0.14, 1.19)$	$0.15(-5.24, 5.55)$	$0.47(-0.28, 1.21)$	$0.00(-4.82, 4.82)$	$-0.00(-5.05, 5.05)$
	PLF	$3.22(-1.88, 8.32)$	$0.67(0.14, 1.19)$	$0.15(-5.24, 5.55)$	$0.47(-0.28, 1.21)$	$0.00(-4.82, 4.82)$	$0.00(-5.05, 5.05)$
2C	EM	$1.89(-5.26, 9.04)$	$0.90(-0.70, 2.51)$	$1.02(-6.24, 8.28)$	$0.45(-1.23, 2.13)$	$0.62(-1.65, 2.89)$	$-0.62(-3.12, 1.88)$
	PLF	$1.95(-5.50, 9.40)$	$0.89(-0.69, 2.46)$	$0.96(-6.60, 8.51)$	$0.46(-1.19, 2.11)$	$0.60(-1.92, 3.12)$	$-0.60(-3.33, 2.13)$
2D	EM	$2.28(0.76, 3.79)$	$1.21(0.66, 1.76)$	$-0.06(-2.08, 1.96)$	$-0.31(-0.97, 0.35)$	$0.00(-1.29, 1.29)$	$-0.00(-1.69, 1.69)$
	PLF	$2.28(0.76, 3.79)$	$1.21(0.66, 1.76)$	$-0.06(-2.08, 1.96)$	$-0.31(-0.97, 0.35)$	$0.00(-1.29, 1.29)$	$0.00(-1.69, 1.69)$
2E	EM	$1.03(-0.50, 2.56)$	$0.78(0.33, 1.23)$	$0.07(-2.01, 2.15)$	$-0.11(-0.67, 0.46)$	$0.00(-1.05, 1.05)$	$-0.00(-1.44, 1.44)$
	PLF	$1.03(-0.50, 2.56)$	$0.78(0.33, 1.23)$	$0.07(-2.01, 2.15)$	$-0.11(-0.67, 0.46)$	$0.00(-1.05, 1.05)$	$0.00(-1.44, 1.44)$
2F	EM	$0.54(-1.61, 2.68)$	$0.70(-0.02, 1.42)$	$0.66(-1.84, 3.16)$	$-0.10(-0.87, 0.67)$	$0.44(-0.27, 1.14)$	$-0.44(-1.61, 0.73)$
	PLF	$0.50(-1.59, 2.59)$	$0.71(-0.00, 1.43)$	$0.69(-1.75, 3.14)$	$-0.11(-0.88, 0.66)$	$0.45(-0.21, 1.11)$	$-0.45(-1.59, 0.69)$
2G	EM	$-1.07(-2.02, -0.11)$	$0.72(0.25, 1.20)$	$0.19(-1.51, 1.89)$	$-0.14(-0.84, 0.56)$	$0.04(-0.16, 0.24)$	$-0.04(-0.46, 0.39)$
	PLF	$-1.09(-2.06, -0.11)$	$0.73(0.25, 1.21)$	$0.21(-1.50, 1.92)$	$-0.15(-0.85, 0.56)$	$0.04(-0.16, 0.24)$	$-0.04(-0.47, 0.38)$
3	EM	$0.78(-0.32, 1.88)$	$0.67(0.37, 0.97)$	$-0.85(-3.57, 1.87)$	$-0.04(-0.91, 0.83)$	$0.00(-0.70, 0.70)$	$0.23(-0.86, 1.33)$
	PLF	$0.78(-0.49, 2.05)$	$0.67(0.35, 0.98)$	$-0.78(-3.75, 2.18)$	$-0.06(-0.96, 0.83)$	$0.00(-0.81, 0.81)$	$0.21(-1.06, 1.48)$
4A	EM	$-0.95(-1.75, -0.15)$	$1.04(0.57, 1.52)$	$0.16(-0.89, 1.21)$	$0.12(-0.54, 0.78)$	$0.00(-0.18, 0.18)$	$-0.00(-0.24, 0.24)$
	PLF	$-0.95(-1.75, -0.15)$	$1.04(0.57, 1.52)$	$0.16(-0.89, 1.21)$	$0.12(-0.54, 0.78)$	$0.00(-0.18, 0.18)$	$0.00(-0.24, 0.24)$
4B	EM	$-0.62(-1.35, 0.12)$	$1.12(0.67, 1.57)$	$0.18(-0.81, 1.17)$	$0.21(-0.47, 0.89)$	$0.00(-0.20, 0.20)$	$-0.00(-0.28, 0.28)$
	PLF	$-0.62(-1.35, 0.12)$	$1.12(0.67, 1.57)$	$0.18(-0.81, 1.17)$	$0.21(-0.47, 0.89)$	$0.00(-0.20, 0.20)$	$0.00(-0.28, 0.28)$
4C	EM	$-1.53(-2.11, -0.96)$	$1.69(1.14, 2.23)$	$0.51(-0.41, 1.44)$	$-0.44(-1.19, 0.31)$	$0.02(-0.03, 0.08)$	$-0.02(-0.17, 0.13)$
	PLF	$-1.54(-2.12, -0.96)$	$1.69(1.15, 2.23)$	$0.52(-0.40, 1.44)$	$-0.44(-1.19, 0.31)$	$0.02(-0.03, 0.08)$	$-0.02(-0.17, 0.13)$
4D	EM	$-1.25(-1.93, -0.57)$	$1.40(0.91, 1.89)$	$0.30(-0.49, 1.09)$	$-0.22(-0.83, 0.39)$	$0.00(-0.11, 0.11)$	$0.02(-0.11, 0.15)$
	PLF	$-1.25(-1.93, -0.57)$	$1.40(0.91, 1.89)$	$0.30(-0.50, 1.09)$	$-0.22(-0.83, 0.39)$	$0.00(-0.11, 0.11)$	$0.02(-0.11, 0.15)$
5A	EM	$1.41(-0.98, 3.80)$	$0.76(0.15, 1.38)$	$-0.10(-3.15, 2.94)$	$-0.08(-0.82, 0.65)$	$0.12(-1.50, 1.74)$	$-0.12(-2.26, 2.03)$
	PLF	$1.41(-1.01, 3.83)$	$0.76(0.14, 1.39)$	$-0.10(-3.17, 2.97)$	$-0.08(-0.82, 0.66)$	$0.12(-1.51, 1.75)$	$-0.12(-2.28, 2.04)$
5B	EM	$0.83(-1.46, 3.13)$	$0.41(0.08, 0.75)$	$-0.24(-3.62, 3.14)$	$0.02(-0.56, 0.60)$	$0.00(-1.56, 1.56)$	$0.13(-1.93, 2.19)$
	PLF	$0.83(-1.47, 3.14)$	$0.41(0.07, 0.75)$	$-0.25(-3.80, 3.29)$	$0.02(-0.59, 0.63)$	$0.00(-1.57, 1.57)$	$0.14(-1.99, 2.26)$
5C	EM	$0.14(-1.79, 2.07)$	$0.44(0.02, 0.87)$	$-2.42(-7.20, 2.35)$	$0.24(-1.56, 2.04)$	$0.00(-0.98, 0.99)$	$0.48(-0.52, 1.48)$
	PLF	$0.15(-1.80, 2.09)$	$0.44(0.01, 0.87)$	$-2.43(-7.21, 2.35)$	$0.24(-1.56, 2.04)$	$0.00(-1.00, 1.00)$	$0.48(-0.53, 1.50)$
5D	EM	$-0.13(-1.68, 1.42)$	$1.25(0.22, 2.28)$	$0.99(-2.21, 4.19)$	$-0.59(-1.85, 0.66)$	$0.45(0.15, 0.76)$	$-0.38(-2.14, 1.38)$
	PLF	$-0.11(-1.66, 1.44)$	$1.24(0.21, 2.27)$	$0.97(-2.20, 4.14)$	$-0.58(-1.83, 0.66)$	$0.45(0.13, 0.77)$	$-0.38(-2.12, 1.36)$
5E	EM	$-0.21(-1.74, 1.33)$	$2.21(0.31, 4.10)$	$1.52(-2.69, 5.73)$	$-1.77(-3.72, 0.18)$	$0.67(0.52, 0.81)$	$-0.63(-3.55, 2.29)$
	PLF	$-0.19(-1.73, 1.35)$	$2.19(0.30, 4.08)$	$1.50(-3.74, 6.75)$	$-1.75(-3.72, 0.23)$	$0.67(0.52, 0.82)$	$-0.63(-4.36, 3.10)$
5F	EM	$-1.33(-25.85, 23.19)$	$0.08(-1.38, 1.54)$	$-0.86(-25.71, 23.99)$	$0.55(-1.63, 2.74)$	$0.13(-4.34, 4.60)$	$0.07(-4.41, 4.55)$
	PLF	$-1.35(-31.95, 29.25)$	$0.08(-1.76, 1.92)$	$-0.79(-31.68, 30.09)$	$0.54(-1.94, 3.01)$	$0.13(-5.34, 5.60)$	$0.06(-5.42, 5.54)$
5G	EM	$-2.38(-6.82, 2.07)$	$0.53(-1.11, 2.17)$	$-4.96(-12.91, 2.99)$	$1.69(-1.47, 4.84)$	$0.05(-0.28, 0.38)$	$0.05(-0.27, 0.38)$
	PLF	$-2.37(-6.79, 2.05)$	$0.53(-1.10, 2.16)$	$-4.98(-12.93, 2.98)$	$1.69(-1.47, 4.85)$	$0.05(-0.28, 0.38)$	$0.05(-0.27, 0.38)$
6A	EM	$-1.22(-1.77, -0.67)$	$1.43(0.96, 1.91)$	$0.11(-1.05, 1.27)$	$-0.17(-1.05, 0.70)$	$0.03(-0.04, 0.11)$	$-0.03(-0.23, 0.16)$
	PLF	$-1.22(-1.78, -0.67)$	$1.44(0.96, 1.91)$	$0.12(-1.05, 1.29)$	$-0.18(-1.06, 0.70)$	$0.04(-0.04, 0.11)$	$-0.04(-0.23, 0.16)$
6B	EM	$-1.41(-2.27, -0.55)$	$1.19(0.65, 1.73)$	$-0.13(-2.02, 1.77)$	$-0.12(-1.21, 0.98)$	$0.00(-0.13, 0.13)$	$-0.00(-0.27, 0.27)$
	PLF	$-1.41(-2.27, -0.55)$	$1.19(0.65, 1.73)$	$-0.13(-2.02, 1.77)$	$-0.12(-1.21, 0.98)$	$0.00(-0.13, 0.13)$	$0.00(-0.27, 0.27)$

Table A4: continued from previous page

Item	Method	b_0	b_1	b_2	b_3	d	d_{DIF}
6C	EM	$-5.21(-9.61, -0.81)$	$2.08(0.00, 4.15)$	$1.52(-3.36, 6.40)$	$-0.74(-3.09, 1.61)$	$0.06(0.02, 0.10)$	$-0.04(-0.10, 0.02)$
	PLF	$-5.21(-9.61, -0.81)$	$2.08(0.00, 4.15)$	$1.55(-3.32, 6.42)$	$-0.75(-3.09, 1.59)$	$0.06(0.02, 0.10)$	$-0.04(-0.10, 0.02)$
6D	EM	$-0.59(-1.79, 0.60)$	$1.14(0.37, 1.92)$	$-0.07(-1.42, 1.28)$	$0.21(-0.75, 1.16)$	$0.07(-0.23, 0.38)$	$-0.06(-0.39, 0.27)$
	PLF	$-0.54(-1.72, 0.65)$	$1.11(0.35, 1.87)$	$-0.12(-1.47, 1.22)$	$0.24(-0.70, 1.18)$	$0.06(-0.26, 0.38)$	$-0.05(-0.39, 0.30)$
6E	EM	$-1.74(-2.84, -0.65)$	$1.03(0.39, 1.66)$	$-0.43(-1.84, 0.98)$	$0.39(-0.48, 1.27)$	$0.09(-0.03, 0.21)$	$-0.03(-0.17, 0.10)$
	PLF	$-1.77(-2.87, -0.67)$	$1.04(0.40, 1.68)$	$-0.41(-1.82, 1.01)$	$0.38(-0.50, 1.26)$	$0.09(-0.03, 0.21)$	$-0.04(-0.17, 0.10)$
6F	EM	$-2.56(-4.12, -1.01)$	$1.77(0.72, 2.81)$	$1.18(-0.59, 2.95)$	$-0.67(-1.83, 0.49)$	$0.16(0.08, 0.24)$	$-0.16(-0.32, -0.01)$
	PLF	$-2.55(-4.11, -1.00)$	$1.76(0.72, 2.80)$	$1.17(-0.60, 2.93)$	$-0.66(-1.82, 0.49)$	$0.16(0.08, 0.24)$	$-0.16(-0.32, -0.01)$
6G	EM	$-2.70(-4.29, -1.11)$	$1.35(0.46, 2.24)$	$0.76(-1.82, 3.33)$	$-0.63(-1.91, 0.66)$	$0.13(0.05, 0.20)$	$-0.10(-0.32, 0.13)$
	PLF	$-2.69(-4.26, -1.11)$	$1.34(0.46, 2.23)$	$0.82(-1.75, 3.38)$	$-0.65(-1.92, 0.62)$	$0.12(0.05, 0.20)$	$-0.10(-0.34, 0.14)$
6H	EM	$-16.03(-35.65, 3.59)$	$7.34(-1.57, 16.26)$	$14.46(-5.18, 34.11)$	$-6.49(-15.42, 2.44)$	$0.23(0.18, 0.27)$	$-0.21(-0.36, -0.06)$
	PLF	$-16.03(-35.65, 3.59)$	$7.34(-1.57, 16.26)$	$14.46(-5.19, 34.11)$	$-6.49(-15.42, 2.44)$	$0.23(0.18, 0.27)$	$-0.21(-0.35, -0.06)$
7A	EM	$-0.01(-1.18, 1.16)$	$0.94(0.31, 1.58)$	$-0.29(-1.81, 1.22)$	$-0.02(-0.83, 0.78)$	$0.01(-0.47, 0.49)$	$-0.01(-0.60, 0.58)$
	PLF	$-0.01(-1.18, 1.17)$	$0.94(0.31, 1.58)$	$-0.29(-1.81, 1.22)$	$-0.02(-0.83, 0.78)$	$0.01(-0.47, 0.49)$	$-0.01(-0.60, 0.58)$
7B	EM	$0.65(-0.92, 2.23)$	$0.52(0.19, 0.84)$	$-1.68(-3.68, 0.31)$	$1.13(0.12, 2.15)$	$0.00(-0.99, 0.99)$	$0.43(-0.57, 1.43)$
	PLF	$0.66(-0.92, 2.24)$	$0.52(0.19, 0.84)$	$-1.68(-3.68, 0.31)$	$1.13(0.12, 2.15)$	$0.00(-0.99, 0.99)$	$0.43(-0.57, 1.43)$
7C	EM	$-0.32(-1.40, 0.76)$	$0.62(0.21, 1.03)$	$0.07(-1.37, 1.51)$	$0.42(-0.31, 1.14)$	$0.00(-0.41, 0.41)$	$0.04(-0.48, 0.56)$
	PLF	$-0.32(-1.39, 0.76)$	$0.62(0.21, 1.03)$	$0.03(-1.40, 1.46)$	$0.44(-0.28, 1.17)$	$0.00(-0.41, 0.41)$	$0.06(-0.45, 0.57)$
7D	EM	$-0.11(-1.18, 0.96)$	$0.90(0.34, 1.47)$	$0.54(-0.73, 1.81)$	$-0.00(-0.68, 0.67)$	$0.12(-0.26, 0.49)$	$-0.12(-0.63, 0.40)$
	PLF	$-0.10(-1.16, 0.96)$	$0.90(0.34, 1.45)$	$0.53(-0.74, 1.79)$	$0.00(-0.66, 0.67)$	$0.11(-0.27, 0.49)$	$-0.11(-0.63, 0.41)$
7E	EM	$-1.32(-2.82, 0.18)$	$0.54(0.02, 1.06)$	$-1.18(-3.22, 0.86)$	$0.82(-0.16, 1.81)$	$0.00(-0.29, 0.29)$	$0.09(-0.21, 0.40)$
	PLF	$-1.31(-2.81, 0.19)$	$0.54(0.02, 1.06)$	$-1.17(-3.21, 0.87)$	$0.82(-0.16, 1.80)$	$0.00(-0.29, 0.29)$	$0.09(-0.21, 0.40)$
7F	EM	$-0.99(-2.43, 0.44)$	$0.45(0.07, 0.82)$	$-0.50(-2.66, 1.67)$	$0.24(-0.58, 1.06)$	$0.00(-0.37, 0.37)$	$0.05(-0.39, 0.50)$
	PLF	$-0.99(-2.43, 0.44)$	$0.45(0.07, 0.82)$	$-0.49(-2.66, 1.69)$	$0.24(-0.59, 1.06)$	$0.00(-0.37, 0.37)$	$0.05(-0.40, 0.50)$