The four-parameter logistic item response model (4PLM) was first mentioned as far back as McDonald (1967) and formally proposed by Barton and Lord (1981). Despite this, for many years, it received very little attention in the literature due to doubts about its utility, as well as technical difficulties related to parameter estimation. Recently, however, the psychometrics community has rekindled its interest in the 4PLM after three decades of neglect. Increasingly, applications of the 4PLM have appeared in the literature as supporting evidence of its usefulness. Several studies argued that the slipping parameter from the 4PLM can be estimated to account for early careless errors to improve ability estimation (Liao et al., 2012; Rulison & Loken, 2009). In psychology, it has been noted that there is value in modeling the lower and upper asymptotes of item response functions (Reise & Waller, 2003; Waller & Reise, 2010) in instances where subjects with higher levels of psychopathology may be reluctant to self-report attitudes, behaviors, and/or experiences which may indicate a social desirability bias (Rouse et al., 1999). In education, recent studies have shown the prevalence and implication of the slipping effect in low-stakes large-scale educational assessments (Culpepper, 2016, 2017).

Despite this work, one daunting challenge regarding the 4PLM pertains to the difficulty in estimating item parameters. The lack of a fast and stable implementation of the 4PLM remains a major barrier to its widespread application in practice. Few studies to date have addressed the item calibration issue for the 4PLM. All the methods fall into two approaches: MCMC methods and marginal maximum likelihood (Feuerstahler & Waller, 2014). For MCMC methods, Loken and Rulison (2010) presented Bayesian model formulations with Metropolis-Hastings (MH) sampling using the general-purpose software (e.g., “bayesmh” module in Stata). More recently, Culpepper (2016) proposed a flexible Gibbs sampling method. Both MCMC methods can provide accurate item estimates, but are inherently time-consuming. Even for Culpepper’s (2016, 2017) Gibbs sampling approach, calibration is expected to take at least 1 hour, which also motivates us to find a faster Bayesian estimation.

Another recent implementation of the 4PLM was done by Waller and Feuerstahler (2017). In their work, they extended the Bayesian modal estimation method (BME) to the 4PLM using the R package mirt (Chalmers, 2012). Compared to MCMC methods, BME is a fast calibration method, but several issues need to be explored to evaluate the utility of BME and its implementation in the mirt package. For example, Guo and Zheng (2019) found that the item parameter estimates yielded by BME were unstable for the 3PLM when changing priors of guessing parameters. Moreover, in the mirt package of version 1.30, the standard errors of all item

---

**Fast Bayesian Estimation for the Four-Parameter Logistic Model (4PLM)**

Chanjin Zheng1*, Shaoyang Guo1*, and Justin L Kern2

**Abstract**
There is a rekindled interest in the four-parameter logistic item response model (4PLM) after three decades of neglect among the psychometrics community. Recent breakthroughs in item calibration include the Gibbs sampler specially made for 4PLM and the Bayes modal estimation (BME) method as implemented in the R package mirt. Unfortunately, the MCMC is often time-consuming, while the BME method suffers from instability due to the prior settings. This paper proposes an alternative BME method, the Bayesian Expectation-Maximization-Maximization-Maximization (BE3M) method, which is developed from by combining an augmented variable formulation of the 4PLM and a mixture model conceptualization of the 3PLM. The simulation shows that the BE3M can produce estimates as accurately as the Gibbs sampling method and as fast as the EM algorithm. A real data example is also provided.

**Keywords**
4PLM, mixture modeling, Gibbs sampler, BME, BE3M

The four-parameter logistic item response model (4PLM) was first mentioned as far back as McDonald (1967) and formally proposed by Barton and Lord (1981). Despite this, for many years, it received very little attention in the literature due to doubts about its utility, as well as technical difficulties related to parameter estimation. Recently, however, the psychometrics community has rekindled its interest in the 4PLM after three decades of neglect.

Increasingly, applications of the 4PLM have appeared in the literature as supporting evidence of its usefulness. Several studies argued that the slipping parameter from the 4PLM can be estimated to account for early careless errors to improve ability estimation (Liao et al., 2012; Rulison & Loken, 2009). In psychology, it has been noted that there is value in modeling the lower and upper asymptotes of item response functions (Reise & Waller, 2003; Waller & Reise, 2010) in instances where subjects with higher levels of psychopathology may be reluctant to self-report attitudes, behaviors, and/or experiences which may indicate a social desirability bias (Rouse et al., 1999). In education, recent studies have shown the prevalence and implication of the slipping effect in low-stakes large-scale educational assessments (Culpepper, 2016, 2017).

Despite this work, one daunting challenge regarding the 4PLM pertains to the difficulty in estimating item parameters. The lack of a fast and stable implementation of the 4PLM remains a major barrier to its widespread application in practice. Few studies to date have addressed the item calibration issue for the 4PLM. All the methods fall into two approaches: MCMC methods and marginal maximum likelihood (Feuerstahler & Waller, 2014). For MCMC methods, Loken and Rulison (2010) presented Bayesian model formulations with Metropolis-Hastings (MH) sampling using the general-purpose software (e.g., “bayesmh” module in Stata). More recently, Culpepper (2016) proposed a flexible Gibbs sampling method. Both MCMC methods can provide accurate item estimates, but are inherently time-consuming. Even for Culpepper’s (2016, 2017) Gibbs sampling approach, calibration is expected to take at least 1 hour, which also motivates us to find a faster Bayesian estimation.

Another recent implementation of the 4PLM was done by Waller and Feuerstahler (2017). In their work, they extended the Bayesian modal estimation method (BME) to the 4PLM using the R package mirt (Chalmers, 2012). Compared to MCMC methods, BME is a fast calibration method, but several issues need to be explored to evaluate the utility of BME and its implementation in the mirt package. For example, Guo and Zheng (2019) found that the item parameter estimates yielded by BME were unstable for the 3PLM when changing priors of guessing parameters. Moreover, in the mirt package of version 1.30, the standard errors of all item

1East China Normal University, Shanghai, China
2University of Illinois Urbana-Champaign, USA
*These authors contributed equally to this work and should be regarded as co-first authors.

**Corresponding Author:**
Shaoyang Guo, Institute of Curriculum & Instruction, Faculty of Education, East China Normal University, Wenke Building, Shanghai 200062, China.
Email: syguo1992@outlook.com

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
parameters cannot be calculated as a consequence of program errors in computing the Hessian matrix when a Beta distribution prior ("expbeta" in mirt) is imposed on guessing or slipping parameters.

At present, we propose an alternative Bayesian modal estimation approach that we term Bayesian Expectation-Maximization-Maximization-Maximization (BE3M) method, which is developed by using a latent variable augmentation approach for the 4PLM in conjunction with the mixture modeling approach for the 3PLM (Zheng et al., 2017). The remainder of the manuscript is as follows: First, we present a detailed derivation of the BME and discuss potential difficulties in item calibration. Second, we present a latent variable augmentation approach to formulating the 4PLM (Béguin & Glas, 2001; Culpepper, 2016), and the resulting BE3M estimation method. Third, we present two simulation studies assessing the accuracy of parameter recovery for the four calibration methods (BE3M, mirt, fourPNO, Stata). Fourth, we apply the four methods to bullying item responses of 7,491 adolescents from the 2005 to 2006 Health Behavior in School-Aged Children (HBSC) study. Lastly, we discuss future research directions and concluding remarks.

**Bayesian Modal Estimation for the 4PLM**

Following Culpepper (2016), the slope-intercept form IRT formulation of the 4PLM is:

\[
P(Y_i = 1, \psi_j) = P_j(\theta_i) = \gamma_j + (1 - \xi_j - \gamma_j)P_j^*(\theta_i),
\]

with

\[
P_j^*(\theta_i) = \frac{1}{1 + \exp(-D(\alpha_j\theta_i - \beta_j))},
\]

where \(Y_i\) is the observed score of examinee \(i\) on item \(j\); \(\alpha_j, \beta_j, \gamma_j, \) and \(\xi_j\) are item slope, intercept, guessing, and slipping parameter of item \(j = 1, 2, 3, ..., J\); \(\theta_i \in \Theta\) is the ability parameter of examinee \(i = 1, 2, 3, ..., N\); \(D\) is a constant, equals 1.702. Let \(\psi_j = (\alpha_j, \beta_j, \gamma_j, \xi_j)^T\) be the \(j\)th column vector of the matrix \(\Psi\) collecting the item parameters for all \(J\) items. Similarly, let \(y\) be the row vector of the matrix \(Y\) collecting all item responses of examinee \(i\). Then, the marginal likelihood function should be

\[
L(Y|\Psi) = \prod_{i=1}^{N} P(y_i|\psi_i)
\]

\[
= \prod_{i=1}^{N} \int_{\Theta} P(y_i|\theta, \Psi)g(\theta|\tau)d\theta
\]

\[
= \prod_{i=1}^{N} \int_{\Theta} \prod_{j=1}^{J} P_j(\theta_i)\left(1 - P_j(\theta_i)\right)^{1-\gamma_j}g(\theta|\tau)d\theta,
\]

with

\[
P(y_i|\theta, \Psi) = \int_{0}^{\infty} P(y_i|\theta, \Psi)g(\theta|\tau)d\theta,
\]

\[
P(y_i|\Psi) = \prod_{j=1}^{J} P_j(\theta_i)\left(1 - P_j(\theta_i)\right)^{1-\gamma_j},
\]

where \(g(\theta_i|\tau)\) is a density function for each examinee \(i\), and \(\tau\) is the vector containing the parameters of the examinee population ability distribution. Please refer to Supplemental Appendix A for a detailed derivation. It is worth pointing out that the BME is a natural extension of the one for 3PLM, but that the Fisher information matrix is slightly larger—it is size \(4 \times 4\) rather than size \(3 \times 3\)—with small diagonal elements for the guessing and slipping parameters. Taken together, we find a great deal of numerical instability in the calculation of the Fisher information matrix, resulting in the BME requiring a much larger sample size than the 3PLM to avoid convergence problems. According to Waller and Feuerstahle (2017), the convergence rate for a sample of 1,000, which is always feasible for 3PLM, is only 9%. The new method in the current paper can address this issue by reformulating the 4PLM as a latent mixture model.

**The Latent Mixture Modeling Reformulation of the 4PLM**

From inspection, we can determine that the 4PLM can be regarded as a latent mixture model. First, note that

\[
P_j(\theta_i) = \gamma_j + (1 - \xi_j - \gamma_j)\times P_j^*(\theta_i)
\]

\[
= P_j^*(\theta_i)\times(1 - \xi_j) + (1 - P_j^*(\theta_i))\times \gamma_j.
\]

Next, we define a latent indicator variable \(W_j\) to indicate whether examinee \(i\) knows the answer to item \(j\), where \(W_j = 1\) if he knows it and 0 if he does not (Béguin & Glas, 2001; Culpepper, 2016). Furthermore, we model \(W_j\) probabilistically as \(W_j \sim Bernoulli(P_j^*(\theta_i))\), meaning that

\[
P(W_j = 1) = P_j^*(\theta_i).
\]

From the mixture-modeling perspective, depending on the value of \(W_j\), the 4PLM can be decomposed into two distinct parts: \(1 - \xi_j\) and \(\gamma_j\). We can further obtain the conditional probabilities \(P(Y_{ij}|W_j = 1, \theta_i, \psi_j)\) as:

\[
P(Y_{ij} = 1|W_j = 1, \theta_i, \psi_j) = 1 - \xi_j,
\]

\[
P(Y_{ij} = 0|W_j = 1, \theta_i, \psi_j) = \xi_j,
\]

\[
P(Y_{ij} = 1|W_j = 0, \theta_i, \psi_j) = \gamma_j,
\]

\[
P(Y_{ij} = 0|W_j = 0, \theta_i, \psi_j) = 1 - \gamma_j.
\]
Thus, we find the joint distribution of $Y_j = 1$ and $W_j$ (conditional on item and person parameters) as follows:

\[
P(Y_j, W_j | \theta_j, \Psi_j) = \begin{cases} 
(1 - \zeta_j)P_j^*(\theta_j), & \text{for } Y_j = 1, \ W_j = 1 \\
\zeta_jP_j^*(\theta_j), & \text{for } Y_j = 0, \ W_j = 1 \\
\gamma_j(1 - P_j^*(\theta_j)), & \text{for } Y_j = 1, \ W_j = 0 \\
(1 - \gamma_j)(1 - P_j^*(\theta_j)), & \text{for } Y_j = 0, \ W_j = 0
\end{cases}
\]  

(8)

Let $w_j$ be the latent indicator vector for examinee $i$. Then the joint distribution for the new augmented complete data $(\mathbf{Y}, \mathbf{W}, \mathbf{\theta})$ is

\[
P(\mathbf{y}, \mathbf{w}, \mathbf{\theta}|\Psi, \tau) = P(\mathbf{y}, \mathbf{w}|\Psi, \mathbf{\theta})g(\theta|\tau)
\]  

(9)

where

\[
P(\mathbf{y}, \mathbf{w}|\Psi, \mathbf{\theta}) = \prod_{j=1}^J \left[\left(1 - \zeta_j\right)P_j^*(\theta_j)\right]^{\gamma_j(1 - P_j^*(\theta_j))} \\
\times \left[\zeta_jP_j^*(\theta_j)\right]^{(1 - \gamma_j)(1 - P_j^*(\theta_j))} \\
\times \left[\left(1 - \gamma_j\right)(1 - P_j^*(\theta_j))\right]^{N(1 - \gamma_j)} \\
\times \left[\left(1 - \gamma_j\right)(1 - P_j^*(\theta_j))\right]^{(1 - \gamma_j)(1 - \gamma_j)}
\]  

(10)

Just like the MMLE/EM algorithm, the BE3M algorithm also marginalizes out the latent ability variable, so the likelihood function of the BE3M without priors is:

\[
L(\mathbf{Y}, \mathbf{W}|\Psi) = \prod_{i=1}^N P(\mathbf{y}_i, \mathbf{w}_i|\Psi)
\]  

(11)

This formulation follows Culpepper (2016) for 4PNO. In contrast to his work, however, we now develop a Bayesian modal method to estimate model parameters for the 4PLM. This approach follows the basic premise of an EM algorithm and proposes to split the maximization step into three parts: first maximizing over $\gamma_j$, then maximizing over $\zeta_j$, and finally jointly maximizing over $\alpha_j$ and $\beta_j$. We further use prior distributions for the item parameters in the maximization step. As such, we call this approach the Bayesian Expectation-Maximization-Maximization-Maximization (BE3M) method.

**Expectation Step and Artificial Data**

In the expectation step, we take the expectation of the log-likelihood over the incomplete portion of the data, the latent variables $W_j$ and $\theta_j$. In the maximization step, we maximize the posteriors for the item parameters. This maximization process is equivalent to finding the zero of the summation of the first derivatives of the expected log-likelihood and the first derivatives of the priors. Thus, here we find the expected value of the first derivatives of the log-likelihood function. To begin, the first derivatives of the log-likelihood for item $j$ is
with
\[ f(\theta|y_i, w_i, \boldsymbol{r}, \Psi) = \frac{P(y_i|\Psi, \theta) g(\theta|\tau)}{\int_0^\infty P(y_i|\Psi, \theta) g(\theta|\tau) d\theta}, \]  

(13)

\[ P(y_i|\Psi, \theta) = \prod_{j=1}^J P_j(\theta_i)^{y_{ij}} [1 - P_j(\theta_i)]^{1-y_{ij}}. \]

(14)

Notice that the above expression includes the latent variable \( W \), so when we take the expectation of the log-likelihood, it will be necessary to calculate \( E(W_i|Y_i, \theta, \Psi) \). To find this, we first find the conditional distribution of \( W \) using Bayes' rule:

\[ P(W_i = 1|Y_i = 1, \theta, \Psi) = \frac{1 - \zeta_j P_j(\theta_i)}{(1 - \zeta_j) P_j^*(\theta_i) + \zeta_j (1 - P_j^*(\theta_i))} \]

(15)

Thus, we find the conditional expectation of \( W_i \) to be

\[ E(W_i|Y_i, \theta, \Psi) = \frac{(1 - \zeta_j) P_j^*(\theta_i)}{P_j^*(\theta_i) + (1 - \zeta_j)(1 - P_j^*(\theta_i))}. \]

(17)

Finally, let \( X_k, k = 1, 2, \ldots, q \), be nodes on the ability scale with associated weights \( A(X_k) \). Using the Hermite-Gauss quadrature method to approximate the integral, the first derivatives of the expected log-likelihood function for each item is:

\[ E \left[ \frac{\partial \ln L(Y, W|\Psi)}{\partial \Psi_j} \right] \approx \sum_{i=1}^N \left[ \sum_{k=1}^q G_i(X_k - \bar{X}_k)|P_j^*(X_k)| \right] \cdot \sum_{k=1}^q \frac{\partial P_j^*(X_k)}{\partial \Psi_j} \]

(18)

with

\[ f(X_k|y_i, w_i, \boldsymbol{r}, \Psi) = f(X_k|y_i, \boldsymbol{r}, \Psi) \cdot A(X_k) / \sum_{k=1}^q P(X_k|Y_i, \theta, \Psi) A(X_k), \]

(19)

\[ P(Y_i|X_k, \Psi) = \prod_{j=1}^J P_j(X_k)^{y_{ij}} [1 - P_j(X_k)]^{1-y_{ij}}, \]

(20)

where \( f(X_k|y_i, w_i, \boldsymbol{r}, \Psi) \) is the posterior density of \( \theta_i \) at quadrature point \( X_k \). Following Bock and Aitkin (1981), we can define several “artificial data” points for our algorithm:

\[ \tilde{f}_k = \sum_{i=1}^N f(X_k|y_i, \boldsymbol{r}, \Psi) \cdot \tilde{r}_{jk} \]

\[ = \sum_{i=1}^N f(X_k|y_i, \boldsymbol{r}, \Psi) \cdot \tilde{r}_{jk}^{(w)} \]

\[ = \sum_{i=1}^N E(W_i|Y_i, X_k, \Psi) f(X_k|y_i, \boldsymbol{r}, \Psi), \]

(21)

\[ \tilde{r}_{jk}^{(w)} = \sum_{i=1}^N E(W_i|Y_i, X_k, \Psi) f(X_k|y_i, \boldsymbol{r}, \Psi). \]

(22)

By taking their expectations, we find that these expressions have very intuitive interpretations:

1. \( E(\tilde{f}_k) = \tilde{f}_k \) is the expected number of persons (out of \( N \)) with ability \( X_k \);
2. \( E(\tilde{r}_{jk}) = P_j(X_k) \tilde{f}_k \) is the expected number of persons with ability \( X_k \) to answer item \( j \) correctly;
3. \( E(\tilde{r}_{jk}^{(w)}) = P_j^*(X_k) \tilde{f}_k \) is the expected number of persons with ability \( X_k \) to know the answer to item \( j \) and also answer it correctly.
4. \( E(\tilde{r}_{jk}^{(w)}) = (1 - \zeta_j) P_j^*(X_k) \tilde{f}_k \) is the expected number of persons with ability \( X_k \) that know the answer to item \( j \) and also answer it correctly.

**Maximization Step 1: \( \gamma_j \)-Parameters**

Let \( \lambda_j \) be the summation of the expected first derivatives of the log-likelihood with respect to \( \gamma_j \) and the first derivatives...
of the prior distribution of $\gamma_j$. Following usual practice (Baker & Kim, 2004), we use a beta prior with hyperparameters $\upsilon^{(s)}_j$ and $\rho^{(s)}_j$. Then,

$$
\lambda_{\gamma_j} = \frac{\partial \ln E(L(\mathbf{Y}, \mathbf{W}|\Psi))}{\partial \gamma_j} \\
+ \frac{\partial \ln g(\gamma_j; \upsilon^{(s)}_j, \rho^{(s)}_j)}{\partial \gamma_j}
$$

(23)

$$
\approx \frac{\upsilon^{(s)}_j - 1}{\gamma_j} - \frac{\rho^{(s)}_j - 1}{1 - \gamma_j}
\approx \sum_{i=1}^{n} \sum_{k=1}^{q} \left[ (Y_{i} - \gamma_j)(1 - E(W_{i,j} | Y_{i,j}, \Psi)) \right] \gamma_j(1 - \gamma_j)
\approx f(X_{i} | y_{i}, w_{i}, \tau, \Psi)
$$

(24)

By setting $\lambda_{\gamma_j}$ equal to 0 and solving for $\gamma_j$, we get a closed-form solution for the guessing parameter:

$$
\gamma_j = \frac{\upsilon^{(s)}_j - 1 + \sum_{k=1}^{q} (\tilde{\eta}_j - \tilde{\eta}_j^{(w)})}{\upsilon^{(s)}_j + \rho^{(s)}_j - 2 + \sum_{k=1}^{q} (\tilde{\eta}_j - \tilde{\eta}_j^{(w)})}.
$$

(25)

Standard errors will require the expected second derivative as well, which is

$$
\lambda_{\gamma_j, \gamma_j} = \frac{\partial^2 \ln E(L(\mathbf{Y}, \mathbf{W}|\Psi))}{\partial \gamma_j^2} \\
+ \frac{\partial^2 \ln g(\gamma_j; \upsilon^{(s)}_j, \rho^{(s)}_j)}{\partial \gamma_j^2}
$$

(26)

Maximization Step 2: $\varsigma_j$-Parameters

Let $\lambda_{\varsigma_j}$ be the summation of the expected first derivatives of the log-likelihood with respect to $\varsigma_j$ and the first derivatives of the prior distribution of $\varsigma_j$. As with the $\gamma_j$-parameters, we use a beta prior with hyperparameters $\upsilon^{(s)}_j$ and $\rho^{(s)}_j$. Then,

$$
\lambda_{\varsigma_j} = \frac{\partial \ln E(L(\mathbf{Y}, \mathbf{W}|\Psi))}{\partial \varsigma_j} \\
+ \frac{\partial \ln g(\varsigma_j; \upsilon^{(s)}_j, \rho^{(s)}_j)}{\partial \varsigma_j}
$$

(27)

Maximization Step 3: $\alpha_j$- and $\beta_j$-Parameters

Let $\lambda_{\alpha_j}$ and $\lambda_{\beta_j}$ be the summation of the expected first derivatives of the log-likelihood and the first derivatives of the prior distribution with respect to $\alpha_j$ and $\beta_j$, respectively. In this maximization step, we first solve the estimates of $\alpha_j$ and then obtain the $\alpha_j$ parameter via $\alpha_j = \exp(\alpha_j^*)$. For both the $\alpha_j$- and $\beta_j$-parameters, we use a normal prior with means $\mu_{\alpha_j}$ and $\mu_{\beta_j}$ and variances $\sigma_{\alpha_j}^2$ and $\sigma_{\beta_j}^2$, respectively. Then,

$$
\lambda_{\alpha_j} = \frac{\partial \ln E(L(\mathbf{Y}, \mathbf{W}|\Psi))}{\partial \alpha_j} \\
+ \frac{\partial \ln g(\alpha_j; \mu_{\alpha_j}, \sigma_{\alpha_j}^2)}{\partial \alpha_j}
$$

(28)

$$
\alpha_j = \frac{\partial \ln E(L(\mathbf{Y}, \mathbf{W}|\Psi))}{\partial \alpha_j} + \frac{\partial \ln g(\alpha_j; \mu_{\alpha_j}, \sigma_{\alpha_j}^2)}{\partial \alpha_j}
$$

(29)
The derivative matrix can be written as

\[
\frac{\partial \ln \left( L(Y, \theta | \phi) \right)}{\partial \beta_j} + \frac{\partial \ln g(\beta_j | \mu_p, \sigma_p^2)}{\partial \beta_j}
\]

\[
= -D \sum_{k=1}^q \left[ \frac{\partial}{\partial \beta_j}(\bar{f}_j^{(w)}) - E\left( \bar{f}_j^{(w)} \right) - \frac{\beta_j - \mu_{\beta_p}}{\sigma_{\beta_p}} \right]
\]

\[
= -D \sum_{k=1}^q \left[ \frac{\partial}{\partial \beta_j}(\bar{f}_j^{(w)}) - \frac{\beta_j - \mu_{\beta_p}}{\sigma_{\beta_p}} \right].
\]  

(30)

Unlike the guessing and slipping parameters, \( \alpha_j^* \) and \( \beta_j \) do not have closed-form solutions for their estimates, so it is necessary to find them numerically. The method used here is the Fisher scoring algorithm, which uses the expected values of the second derivatives, along with the first derivatives. These will also be used later for standard error calculations. We find the second derivatives as follows:

\[
\lambda_{\alpha_j^*} = -D^2 \exp \left( 2\alpha_j^* \right) \sum_{k=1}^q \left\{ X_k^P \left( X_k \right) \left[ 1 - P_j^* \left( X_k \right) \right] \bar{f}_k \right\} - \frac{1}{\sigma_{\alpha_j^*}}.
\]

(31)

\[
\lambda_{\beta_j} = -D^2 \sum_{k=1}^q \left\{ P_j^* \left( X_k \right) \left[ 1 - P_j^* \left( X_k \right) \right] \bar{f}_k \right\} - \frac{1}{\sigma_{\beta_j}}.
\]

(32)

and

\[
\lambda_{\alpha_j^*} = D^2 \exp \left( \alpha_j^* \right) \sum_{k=1}^q \left\{ X_k^P \left( X_k \right) \left[ 1 - P_j^* \left( X_k \right) \right] \bar{f}_k \right\}.
\]

(33)

Standard Errors of Parameter Estimation in BE3M

An important index of estimation quality is the standard error (SE). However, one major criticism of the EM algorithm is that parameter estimate SEs are not a natural product of the algorithm, and so other methods have to be devised (McLachlan & Krishnan, 2007). BE3M falls prey to this criticism, just as all members of the EM algorithm family do.

It can be shown that the inverse of the negative expected value of the matrix of second derivatives of the log-likelihood (i.e., the inverse of the information matrix) is the asymptotic covariance matrix of the estimates (Stuart & Kendall, 1968). As such, the square roots of the diagonal elements of the resulting matrix are the asymptotic standard errors of the parameters. Generically, the expected second derivative matrix can be written as

\[
\Lambda = \begin{pmatrix}
\lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} \\
\lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} \\
\lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} \\
\lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} \\
\lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & \lambda_{\alpha_j^*}
\end{pmatrix}
\]

(34)

in which each entry is the expected partial second derivative of the log-likelihood with respect to the subscript parameters for a given item \( j \). Convergence issues resulting from the ill-conditioning of the matrix plague the MMLE/EM for 4PLM just as it does for the 3PLM (Baker & Kim, 2004).

In contrast, due to the divide-and-conquer strategy implemented in BE3M, we contend that the resulting expected second derivative matrix does not have these issues. In BE3M, the item parameter estimation problem is separated into three smaller distinct estimation problems: (1) estimation of the guessing parameter, (2) estimation of the slipping parameter, and (3) joint estimation of the difficulty and discrimination parameters. This separation has the benefit that any instabilities present in one of the three steps will not negatively affect either of the other two steps. Statistically, this separation implies that the covariances between \( \tilde{\gamma}_j \) and \( \tilde{\alpha}_j \), \( \tilde{\beta}_j \), and \( \tilde{\varsigma}_j \) are zero; similarly, the covariances between \( \tilde{\varsigma}_j \) and \( \tilde{\alpha}_j \), \( \tilde{\beta}_j \), and \( \tilde{\gamma}_j \) are zero. Collected together, we find the item \( j \) expected second derivative matrix to be

\[
\Lambda = \begin{pmatrix}
\lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & 0 & 0 \\
\lambda_{\alpha_j^*} & \lambda_{\alpha_j^*} & 0 & 0 \\
0 & 0 & \lambda_{\gamma_j, \gamma_j} & 0 \\
0 & 0 & 0 & \lambda_{\varsigma_j, \varsigma_j}
\end{pmatrix}
\]

(35)

The standard errors of the estimates for item \( j \), then, are the square roots of the diagonal elements of the matrix \( \Sigma = -\Lambda^{-1} \). Furthermore, the standard error of the \( \alpha_j \) parameter can be transformed from \( \Lambda \) via the delta method (Oehlert, 1992).

Simulation Studies

This section presents Monte Carlo results regarding the accuracy of the developed BE3M procedure for estimating 4PLM item parameters, compared to the BME in mirt (version 1.30) package (Chalmers, 2012) and the MCMC in R package fourPLN (Culpepper, 2016) and “bayesmh” module of Stata/SE 16. Results from two simulation studies are reported. The goal of the simulation studies was to assess the influence of various design features, including the location of item threshold, sample size, and priors in educational (Simulation 1) and psychological (Simulation 2) scenarios.

Item and Person Parameter Generation

For Simulation 1, the item parameters were generated in the same way as in Culpepper (2016). Specifically, item parameters
for 20 items were generated from $\alpha_j \sim N(2,0.5)I(\alpha > 0)$, $\beta_j \sim N(0,0.5)$, $\gamma_j \sim Beta(2,8)$, and $\zeta_j \sim Beta(0,1)$ $I(\zeta_j < 1 - \gamma_j)$. For Simulation 2, we used the item parameters for the 23 items calibrated in Waller and Reise (2010) application to an adolescent self-esteem scale; these were also used in Culpepper (2016). For both testing scenarios, three sample sizes (2,500, 5,000, and 10,000) of person parameters were sampled from the standard normal distribution.

The Setting of MCMC in R Package FourPNO

Since Culpepper (2016) introduced an extra latent variable and imposed various truncation restrictions on the item parameters, the priors in the fourPNO package have obvious differences from the traditional MCMC procedures (see Culpepper (2016) for details). Thus, the current studies used the same setting in the fourPNO package as the simulation studies of Culpepper (2016) in each condition: (1) the $\alpha_j$ parameters were sampled from the truncated normal distribution $N(0,2)I(\alpha > 0)$, which implied $E(\alpha_j) = \frac{2}{\sqrt{\pi}}$ in the priors; (2) the $\beta_j$ parameters were sampled from $N(0,2)$; (3) the uniform truncated $Beta(\upsilon, \rho)$ priors were employed for both $\gamma$ and $\zeta$ parameters ($\upsilon_j^i = \rho_j^i = \upsilon_j = \rho_j = 1$); and (4) the $\theta_j$ parameters were sampled from $N(0,1)$.

Furthermore, since Culpepper (2016) concluded that "no more than 50,000 iterations were needed to achieve $R < 1.1$ for all item parameters," only one chain (100,000 MCMC iterations and discarded the first 50,000 as burn-in) were run in the current studies.

The Setting of BE3M, BME in R Package mirt and MCMC in Stata

Both the BE3M and mirt used 61 quadrature points to approximate the Gaussian distribution from −6 to +6 in all simulation studies and real-data examples. The priors for $\alpha$ and $\beta$ parameters were set as follows: $In \alpha \sim N(0,0.25)$ and $\beta \sim N(0,2)$. Furthermore, two different priors for $\gamma$ and $\zeta$ were used in the BE3M, mirt, and Stata:

1. $\gamma, \zeta \sim Beta(2,8)$, which is the default setting for the guessing parameter in the advanced IRT software flexmirt (Houts & Cai, 2015), and
2. $\gamma, \zeta \sim Beta(4,16)$, which is the default setting for the guessing parameter in the classical IRT software BILOG-MG3 (Zimowski et al., 2003)

In sum, for both the educational and psychological testing scenarios, 21 conditions were simulated: Seven estimation method conditions (fourPNO, BE3M with prior conditions 1 and 2, mirt with prior conditions 1 and 2, and Stata with prior conditions 1 and 2) each crossed with three examinee sample size conditions (2,500, 5,000, and 10,000). For each condition, 100 replications were run to reduce sampling error. Each condition was assessed by its accuracy in recovering individual item parameter as defined by bias and the root mean squared error (RMSE) across the 100 replications:

$$bias_j = \frac{\sum^{s=100}i=1(\hat{\psi} - \psi)}{S},$$ (36)

$$RMSE_j = \sqrt{\frac{\sum^{s=100}i=1(\hat{\psi} - \psi)^2}{S}},$$ (37)

Furthermore, since Waller and Feuerstahler (2017) pointed out that the item parameter recovery for the individual parameter cannot reflect the entire situation of the accuracy of fitting the 4PLM, the current studies also introduced the root integrated mean squared error (RIMSE; Ramsay, 1991) to assess the entire accuracy via the item response function of the 4PLM:

$$RIMSE_j = \frac{\sum^{s=100}i=1\int[p_j(\theta) - P_j(\theta)]^2g(\theta)d\theta}{S}.$$ (38)

Results

The results of two simulation studies will be respectively reported at two levels: the individual level for each type of parameters and the whole level for the item response function of the 4PLM.

As for the individual level, the detailed results for two simulation studies are presented in tables in Supplemental Appendices B and C, respectively. Only RIMSEs (Tables 1 and 2) for the condition of 5,000 examinees are summarized and presented here since those for other conditions are very similar. The results of two simulation studies can be briefly summarized as follows: (1) In general, the results of MCMC in the fourPNO package replicate the previous study in both scenarios; (2) The item parameter recovery of BE3M are very similar with MCMC although the RMSEs of the $\alpha$ parameters yielded by the fourPNO package and Stata are slightly larger than those from the BE3M in the first simulation study; (3) The mirt package and the BE3M have similar bias and RMSEs for $\alpha$ and $\beta$ parameters, but the BE3M has better performance in estimating the $\gamma$ and $\zeta$ parameters; and (4) Compared with the mirt package, the BE3M, and Stata can provide more stable estimates when changing priors for $\gamma$ and $\zeta$ parameters.

As for the whole level, the RIMSEs for two simulation studies are presented in Figure 1. In general, with the increasing of the sample size, the RIMSEs of BE3M, fourPNO, Stata, and mirt package gradually decrease, and all of the RIMSEs across different conditions are relatively small. Furthermore, the results of RIMSEs show that there is no
obvious difference between BE3M and BME in R package mirt although the BE3M has better performance in estimating the asymptotes, which is consistent with Waller and Feuerstahler (2017).

### Real Data Example

The 4PLM is applied to bullying items collected as part of the 2005 to 2006 Health Behavior in School-Aged Children study (Iannotti, 2005) funded by the National Institute of Child Health and Human Development. This real data example was chosen to replicate previous results (Culpepper, 2016), and to demonstrate comparability between BE3M and the MCMC approaches (fourPNO package and “bayesmh” module in Stata) in terms of estimation accuracy and BE3M’s advantage in terms of estimation time.

The prior settings for the “bayesmh” module in Stata are the same as those for BE3M and BME in the fourPNO package.
To further compare BE3M with BME, items were also calibrated with BME in the mirt package. In addition, as we have discussed in the introduction, there is no available SEs for all item parameters in the mirt package because the Beta priors (“expbeta” in mirt) were imposed on the guessing and slipping parameters in this example.

The data includes student responses on ten items related to bullying behavior for 7,491 adolescents. The ten bullying items asked students “How often have you bullied another

| Item | True values | RMSEs for $\alpha$ | RMSEs for $\beta$ |
|------|-------------|---------------------|---------------------|
|      |             | Four PNO (2, 8) BE3M (2, 8) mirt (2, 8) Stata (2, 8) | Four PNO (2, 8) BE3M (2, 8) mirt (2, 8) Stata (2, 8) |
| 1    | 1.91 .53    | .34 .29 .26 .30 .28 | .30 .30 .30 .30 .30 |
| 2    | 1.95 .31    | .34 .28 .29 .30 .33 | .33 .33 .33 .33 .33 |
| 3    | 1.50 .07    | .22 .23 .27 .23 .27 | .25 .25 .25 .25 .25 |
| 4    | 1.12 .07    | .16 .20 .24 .20 .23 | .14 .14 .14 .14 .14 |
| 5    | .89 .40     | .14 .16 .20 .15 .18 | .22 .22 .22 .22 .22 |
| 6    | 1.08 .54    | .23 .20 .23 .20 .22 | .34 .34 .34 .34 .34 |
| 7    | 1.16 .55    | .24 .19 .20 .19 .20 | .15 .15 .15 .15 .15 |
| 8    | 1.10 .01    | .18 .21 .24 .20 .23 | .27 .27 .27 .27 .27 |
| 9    | .78 .35     | .24 .16 .12 .17 .14 | .14 .14 .14 .14 .14 |
| 10   | 1.23 .23    | .14 .18 .26 .18 .26 | .21 .21 .21 .21 .21 |
| 11   | 1.34 .55    | .12 .17 .21 .16 .21 | .24 .24 .24 .24 .24 |
| 12   | 1.54 .74    | .30 .24 .24 .26 .26 | .36 .36 .36 .36 .36 |
| 13   | 1.16 .21    | .31 .26 .25 .28 .29 | .21 .21 .21 .21 .21 |
| 14   | .84 .60     | .16 .15 .16 .14 .15 | .15 .15 .15 .15 .15 |
| 15   | 1.13 .17    | .24 .18 .19 .18 .19 | .23 .23 .23 .23 .23 |
| 16   | .79 .94     | .26 .13 .11 .12 .08 | .14 .14 .14 .14 .14 |
| 17   | 1.27 .61    | .16 .16 .23 .16 .22 | .24 .24 .24 .24 .24 |
| 18   | .94 .12     | .26 .14 .17 .13 .15 | .17 .17 .17 .17 .17 |
| 19   | .84 .12     | .23 .10 .13 .08 .10 | .14 .14 .14 .14 .14 |
| 20   | 1.14 .73    | .25 .12 .17 .12 .15 | .13 .13 .13 .13 .13 |
| 21   | 1.10 .28    | .12 .17 .25 .16 .24 | .20 .20 .20 .20 .20 |
| 22   | .72 .38     | .22 .17 .19 .15 .16 | .14 .14 .14 .14 .14 |
| 23   | .88 .37     | .30 .13 .15 .10 .12 | .12 .12 .12 .12 .12 |

Note. Bold indicates the relatively larger RMSEs in the same conditions.
student(s) at school in the past couple of months in the ways listed below?” The items were dichotomized as “1 → 0” and “>1→1” to indicate the student has not bullied or has bullied, respectively, in the given way over the past couple of months. For this application, the latent variable $W_{ij}$ can be interpreted as

$$W_{ij} = \begin{cases} 1, & \text{Student } i \text{ bullied another student as asked in item } j, \\ 0, & \text{Student } i \text{ did not bully another student as asked in item } j. \end{cases}$$

so that $\varsigma_j$ is the probability that a student who bullied over the past couple of months neglected to report bullying. On the other hand, it can be expected for $\gamma_j$ to be zero in this application, because it is unlikely for students who did not bully to report bullying behaviors (Culpepper, 2016).

**Results**

Item calibration results from the three methods are summarized in Table 3. The estimates and SEs yielded by fourPNO are the same as in Culpepper’s (2016) study due to we used identical package (fourPNO) and estimation conditions. The BE3M produces point estimates comparable to both MCMC approaches, and estimates of the BE3M and Stata remain stable when changing priors. Unlike the performance of the mirt package in the simulation studies, it is interesting to note that the changing of priors has no obvious impact on the guessing and slipping parameters, but rather on the sloping parameters for the Item 1 and 2, which perhaps caused by the unstable Hessian matrix of the BME method. The implication is that careful consideration of priors should be taken if one uses the mirt package (which implements the BME method) to pursue precise estimates of the 4PLM.

In terms of SEs, the two priors do not seem to influence the quality of estimation with similar values for the SEs in general. We also find that the SEs for the MCMC approaches are similar to those obtained with BE3M, with the exceptions of the sloping parameters.

The main finding is that BE3M is much more computationally efficient than the MCMC. The Gibbs sampler in the fourPNO package required approximately 47 minutes to complete 100,000 iterations and the MH sampler in Stata required approximately 3 hours to complete 10,000 iterations with $N=7,491$ using a 2.4 GHz processor and 6 GB of RAM while BE3M only needed about 1 minute. We hope that this finding in particular will help to promote further applications of the 4PLM in practice.

**Discussion**

In recent years, interest in the four-parameter IRT model has been on the rise in measurement research. The current study presented a new Bayesian formulation of the 4PLM model. Through two simulation studies and a real data analysis, we demonstrated that the BE3M combines the strengths of these two approaches; the BE3M can produce estimates as accurately as the Gibbs sampling method as quickly as the Bayes model estimation approach using mirt. To facilitate
### Table 3. The Estimated Item Parameters and the SEs for HBSC Data.

| Item | $\alpha$ (SE) | $\beta$ (SE) | $\beta$ (SE) |
|------|---------------|--------------|--------------|
|      | Stata | BE3M | mirt | Stata | BE3M | Mirt |
| 1    | 4.44 (.44) | 6.16 (.72) | 6.31 (.79) | 5.20 (.61) | 5.40 (.51) | 5.75 | 6.32 |
| 2    | 3.54 (.33) | 3.59 (.52) | 3.67 (.63) | 3.54 (.10) | 3.70 (.11) | 3.62 | 4.14 |
| 3    | 1.21 (.07) | 1.31 (.09) | 1.39 (.09) | 1.25 (.03) | 1.36 (.04) | 1.25 | 1.35 |
| 4    | 1.41 (.06) | 1.49 (.06) | 1.55 (.09) | 1.50 (.04) | 1.56 (.05) | 1.49 | 1.55 |
| 5    | 1.63 (.09) | 1.66 (.11) | 1.74 (.10) | 1.71 (.06) | 1.79 (.06) | 1.68 | 1.75 |
| 6    | 1.93 (.10) | 2.02 (.12) | 2.06 (.11) | 2.09 (.07) | 2.16 (.08) | 2.04 | 2.12 |
| 7    | 2.44 (.15) | 2.50 (.15) | 2.58 (.14) | 2.78 (.12) | 2.91 (.13) | 2.66 | 2.76 |
| 8    | 1.58 (.08) | 1.63 (.10) | 1.68 (.09) | 1.65 (.05) | 1.72 (.05) | 1.63 | 1.70 |
| 9    | 2.23 (.15) | 2.26 (.11) | 2.36 (.16) | 2.49 (.10) | 2.59 (.11) | 2.36 | 2.45 |
| 10   | 2.29 (.16) | 2.41 (.15) | 2.44 (.19) | 2.58 (.11) | 2.70 (.12) | 2.44 | 2.54 |

Note. Bold indicates an obvious difference in estimates when changing priors.
calibration, the authors have made available an R package IRTBEMM (will be published on CRAN soon). The authors believe this package offers a better alternative for 4PLM calibration, helping to promote the use of the 4PLM in educational and psychological research.

Several opportunities exist for future research related to the 4PLM. First, future research can help to identify other substantive research areas and large survey datasets that could benefit from modeling the slipping effect (Beghetto, 2019). Second, for successful, high-efficient integration into a CAT testing system, being able to implement online calibration of item parameters is important (Zheng, 2016). Since the method proposed at present is a member of the EM family to which most online calibration algorithms belong, there is little barrier to developing online calibration methods based on the BE3M.

Author’s Note
Chanjin Zheng and Shaoyang Guo contributed to the work equally and should be regarded as co-first authors. Chanjin Zheng is affiliated with the Department of Educational Psychology, Faculty of Education, East China Normal University, China. Shaoyang Guo is affiliated with the Institute of Curriculum & Instruction, Faculty of Education, East China Normal University, China.

Ethical Approval
This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent
This article does not contain any studies with human participants performed by any of the authors.

Data Availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially supported by the Flower of Happiness Project in social science of East China Normal University (2020ECNU-XFZH007, 2019ECNU-XFZH015) and the Peak Discipline Construction Project of Education at East China Normal University.

ORCID iDs
Chanjin Zheng https://orcid.org/0000-0001-7771-0543
Shaoyang Guo https://orcid.org/0000-0002-2480-8800

Supplemental Material
Supplemental material for this article is available online.

References
Baker, F. B., & Kim, S.-H. (2004). Item response theory: Parameter estimation techniques (2nd ed.). CRC Press.
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. https://doi.org/10.1002/j.2333-8504.1981.tb01255.x
Beghetto, R. A. (2019). Large-scale assessments, personalized learning, and creativity: Paradoxes and possibilities. *ECNU Review of Education, 2*(3), 311–327. https://doi.org/10.1177/2096531119879863
Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika, 66*(4), 541–561. https://doi.org/10.1007/bf0296195
Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459. https://doi.org/10.1007/bf02293801
Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06
Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika, 81*(4), 1142–1163. https://doi.org/10.1007/s11336-015-9477-6
Culpepper, S. A. (2017). The prevalence and implications of slipping on low-stakes, large-scale assessments. *Journal of Educational and Behavioral Statistics, 42*(6), 706–725. https://doi.org/10.3102/1076998617705653
Feuerstahler, L. M., & Waller, N. G. (2014). Estimation of the 4-parameter model with marginal maximum likelihood. *Multivariate Behavioral Research, 49*(3), 285–285. https://doi.org/10.1080/00273171.2014.912889
Guo, S., & Zheng, C. (2019). The Bayesian expectation-maximization-maximization for the 3PLM. *Frontiers in Psychology, 10*, 1175. https://doi.org/10.3389/fpsyg.2019.01175
Houts, C. R., & Cai, L. (2015). flexmirt: Flexible multilevel multidimensional item analysis and test scoring user’s manual version 3.5 RC. North Carolina, CA. https://vpgcentral.com/software/flexmirt/
Iannotti, R. (2005). *Health behavior in school-aged children HBSC, 2003–2006*. MI Inter-University Consortium for Political and Social Research.
Liao, W.-W., Ho, R.-G., Yen, Y.-C., & Cheng, H.-C. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses. *Social Behavior and Personality: An International Journal, 40*(10), 1679–1694. https://doi.org/10.2224/sbp.2012.40.10.1679
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. https://doi.org/10.1111/j.2044-8317.2008.01063.x
MacDonald, R. P. (1967). *The three-parameter logistic model*. New York: Holt, Rinehart and Winston.
McLachlan, G., & Krishnan, T. (2007). *The EM algorithm and extensions* (Vol. 382). John Wiley & Sons.
Oehlert, G. W. (1992). A note on the delta method. *The American Statistician, 46*(1), 27–29.
Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika, 56*(4), 611–630. https://doi.org/10.1007/BF02294494

Reise, S. P., & Waller, N. G. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods, 8*(2), 164–184. https://doi.org/10.1037/1082-989X.8.2.164

Rouse, S. V., Finger, M. S., & Butcher, J. N. (1999). Advances in clinical personality measurement: An item response theory analysis of the MMPI-2 PSY-5 scales. *Journal of Personality Assessment, 72*, 282–307. https://doi.org/10.1207/S15327752JP720212

Rulison, K. L., & Loken, E. (2009). I’ve fallen and I can’t get up: Can high-ability students recover from early mistakes in CAT? *Applied Psychological Measurement, 33*(2), 83–101. https://doi.org/10.1177/0146621608324023

Stuart, A., & Kendall, M. G. (1968). *The advanced theory of statistics*. Macmillan.

Waller, N. G., & Feuerstahler, L. (2017). Bayesian modal estimation of the four-parameter item response model in real, realistic, and idealized data sets. *Multivariate Behavioral Research, 52*(3), 350–370. https://doi.org/10.1080/00273171.2017.1292893

Waller, N. G., & Reise, S. P. (2010). Measuring psychopathology with nonstandard item response theory models: Fitting the four-parameter model to the Minnesota Multiphasic Personality Inventory. In S. Embretson (Ed.), *Measuring psychological constructs: Advances in model based approaches* (pp. 147–173). American Psychological Association.

Zheng, C., Meng, X., Guo, S., & Liu, Z. (2017). Expectation-maximization-maximization: A feasible MLE algorithm for the three-parameter logistic model based on a mixture modeling reformulation. *Frontiers in Psychology, 8*, 2302. https://doi.org/10.3389/fpsyg.2017.02302

Zheng, Y. (2016). Online calibration of polytomous items under the generalized partial credit model. *Applied Psychological Measurement, 40*(6), 434–450. https://doi.org/10.1177/0146621616650406

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *BILOG-MG3 user’s guide*. Scientific Software International.