frontiers | Frontiers in Psychology

Bayesian Analysis of Aberrant Response and Response Time Data

Zhaoyuan Zhang[1,2], Jiwei Zhang[3]* and Jing Lu[4]*

[1] School of Mathematics and Statistics, Yili Normal University, Yining, China, [2] Institute of Applied Mathematics, Yili Normal University, Yining, China, [3] School of Mathematics and Statistics, Yunnan University, Kunming, China, [4] Key Laboratory of Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, China

In this article, a highly effective Bayesian sampling algorithm based on auxiliary variables is proposed to analyze aberrant response and response time data. The new algorithm not only avoids the calculation of multidimensional integrals by the marginal maximum likelihood method but also overcomes the dependence of the traditional Metropolis–Hastings algorithm on the tuning parameter in terms of acceptance probability. A simulation study shows that the new algorithm is accurate for parameter estimation under simulation conditions with different numbers of examinees, items, and speededness levels. Based on the sampling results, the powers of the two proposed Bayesian assessment criteria are tested in the simulation study. Finally, a detailed analysis of a high-state and large-scale computerized adaptive test dataset is carried out to illustrate the proposed methodology.

Keywords: aberrant responses, Bayesian inference, mixture hierarchical model, Pólya-gamma distribution, rapid guessing behavior, Gibbs sampling algorithm

1. INTRODUCTION

In educational psychological assessments, examinees often perform different types of test-taking behaviors (Bolt et al., 2002; Boughton and Yamamoto, 2007; Goegebeur et al., 2008; Chang et al., 2014; Wang and Xu, 2015; Wang et al., 2018; Man et al., 2018; Man and Harring, 2021). One is the solution behavior, in which the examinee gives a response after careful consideration to each part of an item (Schnipke and Scrams, 1997; Bolt et al., 2002; Wise and Kong, 2005; Wang and Xu, 2015). An alternative is the rapid guessing behavior, in which the examinee simply seeks to obtain an answer quickly without a deep thought process; this behavior often occurs in high-stakes tests owing to insufficient time and in low-stakes tests owing to lack of motivation. In fact, the traditional item response theory (IRT) model is based on the assumption that the correct response probability increases with the ability of the test taker under the solution behavior. The correct response probability under the rapid guessing behavior is actually rarely dependent on the measure constructed by the test (Lord and Novick, 1968; Wise and DeMars, 2006; Boughton and Yamamoto, 2007; Goegebeur et al., 2008). Numerous studies have shown that the presence of rapid guessing behavior inevitably leads to biased inferences of the item and person parameters (Bolt et al., 2002; Wise and DeMars, 2006; Boughton and Yamamoto, 2007; Goegebeur et al., 2008; Chang et al., 2014; Wang and Xu, 2015; Wang et al., 2018). Therefore, appropriate models need to be constructed to capture both solution behavior and rapid guessing behavior to reduce these biased parameter estimates. Before we analyze aberrant response behavior, we provide an explanation of the change point, which is the cut-off point at which an examinee adopts different response strategies. By considering a change point, Bolt et al. (2002) classified examinees in the speeded group before the

change point and found that they were more likely to adopt the solution behavior, whereas examinees who transferred from the speeded group to the non-speeded group after the change point were more likely to choose the rapid guessing behavior. In contrast to models using fixed change point locations, Boughton and Yamamoto (2007) proposed the more flexible HYBRID model, which allowed different examinees to have change points at different locations. The model assumes that examinees' responses follow a Rasch model until a particular point in a given examinee's test, after which the responses to all items are randomly guessed. Goegebeur et al. (2008) proposed a speeded model with one change point to characterize the gradual switch between response strategies by introducing an additional examinee-specific change-rate parameter. In addition, Wise and DeMars (2006) proposed an effort-moderated IRT model to decompose the correct response probability into a mixture of two sub-models. The two sub-models were used to characterize the solution behavior and rapid guessing behavior, respectively.

In parallel with the abovementioned item response data, response time, which is an important type of important auxiliary information, has been widely used to distinguish between two different behaviors (Schnipke and Scrams, 1997; Wise and DeMars, 2006; van der Linden and Guo, 2008; Wang and Xu, 2015). van der Linden and Guo (2008) found that examinees' response times in a high-stakes achievement test showed a mixture of two different distributions. Similarly, Schnipke and Scrams (1997) verified that the distribution of response times for end-of-test items showed a bimodal pattern in a high-stakes exam. In the study of (Schnipke and Scrams, 1997), a two-state mixture model was proposed to decompose the distribution of response times for each item into two parts. The two parts of the response times quantified the solution behavior and the rapid guessing behavior, respectively. Wang and Xu (2015) proposed a mixture model to consider differences between item responses and response time patterns resulting from the solution behavior and rapid guessing behavior. The mixture model used both item response and response time information and considered multiple switch points for each examinee.

A variety of estimation methods have been proposed to estimate the parameters of the IRT and response time models. In the frequentist framework, the most common method is the marginal maximum likelihood estimation (MMLE) *via* expectation maximization algorithm (Bock and Aitkin, 1981; Baker and Kim, 2004). However, the main drawback of marginal maximum likelihood methods is the inevitable need for tedious approximation of the multidimensional integral using numerical integration (Bock and Schilling, 1997; Rabe-Hesketh et al., 2002, 2005) or Monte Carlo integration (Kuk, 1999; Skaug, 2002) when the latent variables are high dimensional. This is because the number of discrete quadrature points required increases exponentially as the number of latent variables increases linearly during the computation (Converse et al., 2021, p. 1465). Although the adaptive quadrature method has been used to deal with the computational deficiency by using a small number of quadrature points, the problem has not been completely solved (Jiang and Templin, 2019). In addition, the comparison method of the MMLE is simplistic; comparison methods other

than the root mean square error of approximation are seldom used (Zhang et al., 2019). Compared with the MMLE method, first, the Bayesian method allows one to update knowledge by using proper informative priors based on previous studies, the posterior distribution being more precise than the likelihood or the prior alone (Jackman, 2009). The incorporation of proper informative priors into the Bayesian analysis can be used to obtain better results in the case of small to moderate sample sizes. In addition, even if weakly informative inaccurate priors are used, the performance of the Bayesian method does not deteriorate. Second, Bayesian estimation does not rely on asymptotic arguments and can give more reliable results for small samples (Lee and Song, 2004; Song and Lee, 2012). Third, another major advantage of Bayesian analysis is the ability to analyze models that are computationally heavy or impossible to estimate with MMLE. These include models with categorical outcomes with many latent variables and, thus, many dimensions of numerical integrations (Asparouhov and Muthén, 2010b; Muthén, 2010).

In the current study, an efficient Pólya–gamma Gibbs sampling algorithm (Polson et al., 2013) in a fully Bayesian framework is proposed to estimate the parameters of the mixture model of Wang and Xu (2015). Compared with traditional Bayesian sampling algorithms, e.g., the Metropolis–Hastings sampling algorithm (Metropolis et al., 1953; Hastings, 1970; Tierney, 1994; Chib and Greenberg, 1995; Chen et al., 2000), Gibbs sampling algorithm (Geman and Geman, 1984; Tanner and Wong, 1987; Gelfand and Smith, 1990; Albert, 1992; Béguin and Glas, 2001; Fox and Glas, 2001), and the advantages of the Pólya–gamma Gibbs sampling algorithm are presented from multiple perspectives. First, the Pólya–gamma Gibbs sampling algorithm avoids retrospective tuning in the Metropolis–Hastings sampling algorithm if we do not know how to choose a proper tuning parameter or if no value for the tuning parameter is appropriate. It always keeps the drawn samples accepted, thereby increasing the sampling efficiency (Zhang et al., 2020). Second, the Pólya–gamma Gibbs sampling algorithm can transform the non-conjugate model into the conjugate model by using augmented auxiliary variables. With the help of the traditional Gibbs sampling algorithm, posterior sampling is easier to implement (Polson et al., 2013). Third, in Bayesian estimation, prior distributions of model parameters and observed data likelihood produce a joint posterior distribution for the model parameters. The prior specifications and prior sensitivity are important aspects of Bayesian inference (Ghosh and A. Ghosh, 2000). In fact, the Pólya–gamma Gibbs sampling algorithm is not sensitive to the specification of the prior distribution. It can still obtain satisfactory results even if the proper or mis-specification priors are adopted (Zhang et al., 2020).

The rest of this article is organized as follows. Section 2 contains an introduction to the mixture hierarchical model and the corresponding identification restrictions. A detailed implementation of the Pólya–gamma Gibbs sampling algorithm is described in Section 3. In Section 4, two simulations focus on the parameter recovery performance of the Bayesian algorithm using the results of the model assessments. In addition, the quality of the Bayesian algorithm is investigated using

high-state and large-scale computerized adaptive test data in Section 5. We conclude the article with a brief discussion in Section 6.

2. MODELS

Following Wang and Xu (2015), the mixture model is used to distinguish solution behavior from rapid guessing behavior. The correct response probability of examinee i on item j is assumed to follow a mixture decomposition

$$P\left(Y_{ij} = 1 \,\middle|\, \eta_{ij}\right) = \left(1 - \eta_{ij}\right) P\left(Y_{ij} = 1 \,\middle|\, \eta_{ij} = 0\right) + \eta_{ij} P\left(Y_{ij} = 1 \,\middle|\, \eta_{ij} = 1\right),$$

where η_{ij} is a latent response behavior indicator variable, $\eta_{ij} = 1$ denotes the case where examinee i answers item j by rapid guessing behavior, and $\eta_{ij} = 0$ denotes the solution behavior. $P\left(Y_{ij} = 1 \,\middle|\, \eta_{ij} = 0\right)$ quantifies the probability of a correct response resulting from the solution behavior, whereas $P\left(Y_{ij} = 1 \,\middle|\, \eta_{ij} = 1\right)$ captures the probability of a correct response with the rapid guessing behavior. We use the two-parameter logistic (2PL; Birnbaum, 1968) model for the solution behavior,

$$P\left(Y_{ij} = 1 \,\middle|\, \eta_{ij} = 0, \theta_i, a_j, b_j\right) = \frac{\exp\left[a_j\left(\theta_i - b_j\right)\right]}{1 + \exp\left[a_j\left(\theta_i - b_j\right)\right]},$$

where a_j and b_j are the discrimination and difficulty parameters of item j, and θ_i denotes the ability of the $i$the examinee. The probability that examinee i answers item j correctly by the rapid guessing behavior is g_j; this is an item-specific probability:

$$P\left(Y_{ij} = 1 \,\middle|\, \eta_{ij} = 1\right) = g_j.$$

In parallel with the mixture item response model, the observed response time T_{ij}^{obs} is

$$T_{ij}^{obs} = \left(1 - \eta_{ij}\right) T_{ij} + \eta_{ij} C_{ij},$$

where T_{ij} represents the time required for examinee i to respond to item j using solution behavior, and C_{ij} represents the time required for examinee i to respond to item j using rapid guessing behavior. Therefore, given latent indicator variable η_{ij}, the density function of observed response time T_{ij}^{obs} can be denoted as

$$p_{ij}\left(t_{ij} \,\middle|\, \eta_{ij}\right) = \left(1 - \eta_{ij}\right) f_{ij}\left(t_{ij}\right) + \eta_{ij} h_{ij}\left(t_{ij}\right),$$

where f and h represent corresponding density functions of T_{ijv} and C_{ijv}.

Response times on test items have been modeled in various families of distributions in psychometric applications, including exponential (Scheiblechner, 1979), gamma (Maris, 1993), Weibull (Rouder et al., 2003), log-normal race (Rouder et al., 2015), and semi-parametric models (Wang et al., 2013). Response time data are non-negative, and their distributions tend to be positively skewed. The log transformation would move

positively skewed distributions toward symmetric shapes. We chose the log-normal distribution (van der Linden, 2006) for response times with solution behavior:

$$\log\left(T_{ij}\right) = \lambda_j - \tau_i + e_{ij}, \ \ e_{ij} \sim N\left(0, \sigma_j^2\right),$$

where λ_j is the time intensity of item j; a higher value of λ_j indicates that the item is expected to consume more time. τ_i is a speed parameter of examinee i; a higher value of τ_i means that the examinee works faster and a lower response time is expected. σ_j^2 allows for differences between the variances of log-times on different items. Following the "common-guessing" (Schnipke and Scrams, 1997), the response times of the guessing behavior have a common log-normal distribution

$$\log\left(C_{ij}\right) \sim N\left(\mu_c, \sigma_c^2\right).$$

To capture across-person relationships between speed and accuracy, we assume that the ability and speed parameters have a bivariate normal distribution, to explore whether examinees with higher ability tend to answer items faster, i.e.,

$$\boldsymbol{\xi}_i = \left(\theta_i, \tau_i\right)' \sim N\left(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P\right),$$

with mean vector

$$\boldsymbol{\mu}_P = \left(\mu_\theta, \mu_\tau\right)'$$

and covariance matrix

$$\boldsymbol{\Sigma}_P = \begin{pmatrix} \sigma_\theta^2 & \sigma_{\theta\tau} \\ \sigma_{\tau\theta} & \sigma_\tau^2 \end{pmatrix}.$$

2.1. Model Identification

In the 2PL model, to eliminate the trade-off between ability θ and difficulty parameter b in location, we only need to fix the mean population level of ability to zero. That is, $\mu_\theta = 0$. To eliminate the trade-off between ability θ and discrimination parameter a in scale, we need to restrict the variance population level of ability to one. That is, $\sigma_\theta^2 = 1$. For the response time model with the solution behavior, to eliminate the trade-off between speed parameter τ and time intensity parameter λ in location, we need to fix the mean population level of speed to zero. That is, $\mu_\tau = 0$.

There are several widely used identification restriction methods for two-parameter IRT models. The identification restrictions of our models are based on the following methods.

(1) Fix the mean population level of ability to zero and the variance population level of ability to one (Lord and Novick, 1968; Bock and Aitkin, 1981; Fox and Glas, 2001; Fox, 2010). That is, $\theta \sim N(0, 1)$.

(2) Restrict the sum of item difficulty parameters to zero and the product of item discrimination parameters to one (Fox and Glas, 2001; Fox, 2005, 2010). That is,

$$\sum_{j=1}^{J} b_j = 0 \quad \text{and} \quad \prod_{j=1}^{J} a_j = 1.$$

(3) Fix the item difficulty parameter to a specific value, most often zero and restrict the discrimination parameter to a specific value, most often one (Fox and Glas, 2001; Fox, 2010). That is, $b_1 = 0$ and $a_1 = 1$.

3. BAYESIAN ESTIMATION USING MCMC SAMPLING

Let $\boldsymbol{\Omega} = (\eta_{ij}, , \theta_i, a_j, b_j, \lambda_j, \tau_i, \sigma_j^2, \mu_a, \sigma_a^2, \mu_b, \sigma_b^2, \mu_\lambda, \sigma_\lambda^2, \mu_c, \sigma_c^2, g_j, \sigma_{\theta\tau}, \sigma_\tau^2, \pi_i)$; then, the full joint posterior person and item parameters given $\boldsymbol{Y}, \boldsymbol{T}$, and $\boldsymbol{\eta}$ is

$L(\boldsymbol{\Omega} \,|\, \boldsymbol{Y}, \boldsymbol{T})$

$$
\begin{aligned}
=& \prod_{i=1}^{N} \prod_{j=1}^{J} \left[\pi_i g_j h\left(t_{ij}; \mu_c, \sigma_c^2\right) \right]^{\eta_{ij} \cdot Y_{ij}} \left[\pi_i \left(1-g_j\right) h\left(t_{ij}; \mu_c, \sigma_c^2\right) \right]^{\eta_{ij} \cdot (1-Y_{ij})} \\
& \times \left[\left(1-\pi_i\right) P\left(Y_{ij}=1 \,|\, \eta_{ij}=0, a_j, b_j, \theta_i\right) f\left(t_{ij}; \lambda_j, \tau_i, \sigma_j^2\right) \right]^{(1-\eta_{ij}) \cdot Y_{ij}} \\
& \times \left[\left(1-\pi_i\right) P\left(Y_{ij}=0 \,|\, \eta_{ij}=0, a_j, b_j, \theta_i\right) f\left(t_{ij}; \lambda_j, \tau_i, \sigma_j^2\right) \right]^{(1-\eta_{ij}) \cdot (1-Y_{ij})} \\
& \times p\left(\theta_i, \tau_i; \boldsymbol{\mu}_P, \Sigma_P\right) p\left(a_j\right) p\left(b_j\right) p\left(\lambda_j\right) p\left(\boldsymbol{\mu}_P, \Sigma_P\right),
\end{aligned}
\tag{1}
$$

where π_i is the probability that examinee i uses the rapid guessing behavior, i.e., $\pi_i = P\left(\eta_{ij}=1\right)$.

3.1. Pólya–Gamma Gibbs Sampling Algorithm

Polson et al. (2013) proposed a new data augmentation strategy for fully Bayesian inference in logistic regression. This data augmentation approach used a new class of Pólya–gamma distribution, in contrast to the data augmentation algorithm of Albert and Chib (1993), which was based on a truncated normal distribution. Here, we introduce the Pólya–gamma distribution. **Definition**: Let $\{B_k\}_{k=1}^{+\infty}$ be an *independent and identically distributed* random variable sequence from a gamma distribution with parameters β and 1. That is, $B_k \sim \text{gamma}\left(\beta, 1\right)$. A random variable W follows a Pólya–gamma distribution with parameters $\beta > 0$ and $\varrho \in R$, denoted $W \sim \text{PG}\left(\beta, \varrho\right)$, if

$$
W \overset{D}{=} \frac{1}{2\pi} \sum_{k=1}^{+\infty} \frac{B_k}{\left(k-\frac{1}{2}\right)^2 + \frac{\varrho^2}{4\pi^2}},
$$

where $\overset{D}{=}$ denotes equality in distribution. In fact, the Pólya–gamma distribution is an infinite mixture of gamma distributions, which provides the ability to sample from gamma distributions.

Based on Theorem 1 of Polson et al. (2013, page 1341, Equation 7), the likelihood contribution of the ith examinee answering the jth item under the solution behavior category $\eta_{ij} =$

0 can be expressed as

$$
\begin{aligned}
L\left(a_j, b_j, \theta_i\right) =& \frac{\left\{\exp\left[a_j\left(\theta_i-b_j\right)\right]\right\}^{Y_{ij}}}{1 + \left\{\exp\left[a_j\left(\theta_i-b_j\right)\right]\right\}} \propto \exp\left\{k_{ij}\left[a_j\left(\theta_i-b_j\right)\right]\right\} \\
& \times \int_0^\infty \exp\left\{-\frac{W_{ij}\left[a_j\left(\theta_i-b_j\right)\right]^2}{2}\right\} p\left(W_{ij} \,|\, 1, 0\right) dW_{ij},
\end{aligned}
\tag{2}
$$

where $k_{ij} = Y_{ij} - \frac{1}{2}$. $p\left(W_{ij} \,|\, 1, 0\right)$ is the conditional density of W_{ij}. That is, $W_{ij} \sim \text{PG}\left(1, 0\right)$. The auxiliary variable W_{ij} follows a Pólya–gamma distribution with parameters $(1, 0)$. Within the solution behavior category $\eta_{ij} = 0$, the full conditional distribution of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\theta}$ given the auxiliary variables, $\boldsymbol{W}$ can be written as

$p\left(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\theta} \,|\, \boldsymbol{\eta}, \boldsymbol{W}, \boldsymbol{Y}\right)$

$$
\begin{aligned}
\propto & \left\{ \prod_{i=1}^{N} \prod_{j=1}^{J} \left[\exp\left\{k_{ij}\left[a_j\left(\theta_i-b_j\right)\right]\right\} \exp\left[-\frac{W_{ij}\left[a_j\left(\theta_i-b_j\right)\right]^2}{2}\right] \right] \right\}^{I(\eta_{ij}=0)} \\
& \times \left\{ \prod_{i=1}^{N} p\left(\theta_i \,|\, \tau_i, \boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P\right) \right\}^{I(\eta_{ij}=0)} \left\{ \prod_{j=1}^{J} \left[p\left(a_j\right) p\left(b_j\right) \right] \right\}^{I(\eta_{ij}=0)},
\end{aligned}
\tag{3}
$$

where $p\left(a_j\right)$ and $p\left(b_j\right)$ are the prior distributions for a_j and b_j. It is known that there are relationships between the latent ability and speed parameter, which can be constructed by a bivariate normal prior distribution $\begin{pmatrix} \theta_i \\ \tau_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\theta \\ \mu_\tau \end{pmatrix}, \boldsymbol{\Sigma}_P\right)$. Therefore, the conditional prior distribution of θ_i is the normal distribution

$$
\theta_i \,|\, \tau_i, \boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P \sim N\left(\mu_{\theta|\tau}, \sigma_{\theta|\tau}^2\right),
$$

where $\mu_{\theta|\tau} = \mu_\theta + \sigma_{\theta\tau} \sigma_\tau^{-2}\left(\tau_i - \mu_\tau\right)$ and $\sigma_{\theta|\tau}^2 = \sigma_\theta^2 - \sigma_{\theta\tau} \sigma_\tau^{-2} \sigma_{\tau\theta}$.

Step 1: Sample the auxiliary variable W_{ij}, within the solution behavior category $\eta_{ij} = 0$, given the item discrimination and difficulty parameters a_j, b_j and the ability θ_i. According to Equation (1), the full conditional posterior distribution of the random auxiliary variable W_{ij} is given by

$$
f\left(W_{ij} \,|\, a_j, b_j, \theta_i\right) \propto \exp\left[-\frac{W_{ij}\left[a_j\left(\theta_i-b_j\right)\right]^2}{2}\right] p\left(W_{ij} \,|\, 1, 0\right).
$$

According to Biane et al. (2001) and Polson et al. (2013; p. 1341), the density function $p\left(W_{ij} \,|\, 1, 0\right)$ can be written as

$$
p\left(W_{ij} \,|\, 1, 0\right) = \sum_{v=0}^{\infty} (-1)^v \frac{(2k+1)}{\sqrt{2\pi W_{ij}}} \exp\left[-\frac{(2k+1)^2}{8 W_{ij}}\right].
$$

Therefore, $f\left(W_{ij} \,|\, a_j, b_j, \theta_i\right)$ is proportional to

$$
\sum_{v=0}^{\infty} (-1)^v \frac{(2k+1)}{\sqrt{2\pi W_{ij}}} \exp\left[-\frac{(2k+1)^2}{8 W_{ij}} - \frac{W_{ij}\left[a_j\left(\theta_i-b_j\right)\right]^2}{2}\right].
$$

Finally, the specific form of the full conditional distribution of W_{ij} is as follows:

$$W_{ij} \sim \text{PG}\left(1, \left|a_j\left(\theta_i - b_j\right)\right|\right).$$

Next, Gibbs samplers are used to draw the item parameters.

Step 2: Sample the discrimination parameter a_j for each item j. The prior distribution of a_j is assumed to follow a truncated normal distribution, i.e., $a_j \sim N\left(\mu_a, \sigma_a^2\right) I\left(a_j > 0\right)$. Given $\boldsymbol{Y}$, $\boldsymbol{W}$, b_j, and $\boldsymbol{\theta}$, the fully conditional posterior distribution of a_j is given by

$$p\left(a_j \mid \boldsymbol{Y}, \boldsymbol{W}, b_j, \boldsymbol{\theta}\right)$$
$$\propto \prod_{i=1}^{N}\left\{\frac{\left\{\exp\left[a_j\left(\theta_i - b_j\right)\right]\right\}^{Y_{ij}}}{1 + \exp\left[a_j\left(\theta_i - b_j\right)\right]} f\left(W_{ij} \mid a_j, b_j, \theta_i\right)\right\} p\left(a_j\right),$$

where $f\left(W_{ij} \mid a_j, b_j, \theta_i\right)$ is given by the following equation (for details, refer to Polson et al., 2013; p. 1341):

$$f\left(W_{ij} \mid a_j, b_j, \theta_i\right) = \left\{\cosh\left(2^{-1}\left|a_j\left(\theta_i - b_j\right)\right|\right)\right\} \frac{2^0}{\Gamma(1)}$$
$$\times \sum_{v=0}^{\infty}(-1)^v \frac{(2k+1)}{\sqrt{2\pi W_{ij}}}$$
$$\times \exp\left[-\frac{(2k+1)^2}{8 W_{ij}} - \frac{W_{ij}\left[a_j\left(\theta_i - b_j\right)\right]^2}{2}\right].$$

After rearrangement, the full conditional posterior distribution of a_j can be written as follows:

$$p\left(a_j \mid \boldsymbol{Y}, \boldsymbol{W}, b_j, \boldsymbol{\theta}\right) \propto \prod_{i=1}^{N}\left\{\frac{\left\{\exp\left[a_j\left(\theta_i - b_j\right)\right]\right\}^{Y_{ij}}}{1 + \exp\left[a_j\left(\theta_i - b_j\right)\right]}\right.$$
$$\times \left[\cosh\left(2^{-1}\left|a_j\left(\theta_i - b_j\right)\right|\right)\right]$$
$$\left. \times \exp\left[-\frac{\left[a_j\left(\theta_i - b_j\right)\right]^2 W_{ij}}{2}\right]\right\} p\left(a_j\right).$$

Therefore, the fully conditional posterior distribution of a_j follows a normal distribution truncated at 0 with mean

$$\text{Var}_{a_j} \times \left(\mu_a \sigma_a^{-2} + \left[\sum_{i=1}^{N} W_{ij}\left(\theta_i - b_j\right)^2\right]\left\{\frac{\left[\sum_{i=1}^{N}\left(1 - 2Y_{ij}\right)\left(\theta_i - b_j\right)\right]}{2\left[\sum_{i=1}^{N} W_{ij}\left(\theta_i - b_j\right)^2\right]}\right\}\right)$$

and variance

$$\text{Var}_{a_j} = \left\{\sigma_a^{-2} + \left[\sum_{i=1}^{N} W_{ij}\left(\theta_i - b_j\right)^2\right]\right\}^{-1}.$$

Step 3: Sample the difficulty parameter b_j for each item j. The prior distribution of b_j is assumed to follow a normal distribution with mean μ_b and σ_b^2. That is, $b_j \sim N\left(\mu_b, \sigma_b^2\right)$. Similarly, given

$\boldsymbol{Y}$, $\boldsymbol{W}$, a_j, and $\boldsymbol{\theta}$, the fully conditional posterior distribution of b_j is given by

$$p\left(b_j \mid \boldsymbol{Y}, \boldsymbol{W}, a_j, \boldsymbol{\theta}\right) \propto \prod_{i=1}^{N}\left\{\frac{\left\{\exp\left[a_j\left(\theta_i - b_j\right)\right]\right\}^{Y_{ij}}}{1 + \exp\left[a_j\left(\theta_i - b_j\right)\right]}\right.$$
$$\times \left[\cosh\left(2^{-1}\left|a_j\left(\theta_i - b_j\right)\right|\right)\right]$$
$$\left. \times \exp\left[-\frac{\left[a_j\left(\theta_i - b_j\right)\right]^2 W_{ij}}{2}\right]\right\} p\left(b_j \mid \mu_b, \sigma_b^2\right).$$

Therefore, the fully conditional posterior distribution of b_j follows a normal distribution with mean

$$\text{Var}_{b_j} \times \left(\mu_b \sigma_b^{-2} + \sum_{i=1}^{N}\left[a_j^2 W_{ij}\right]\left(\frac{\sum_{i=1}^{N}\left(2a_j^2 \theta_i W_{ij} - 2Y_{ij}a_j + a_j\right)}{2\sum_{i=1}^{N}\left[a_j^2 W_{ij}\right]}\right)\right)$$

and variance

$$\text{Var}_{b_j} = \left\{\sigma_b^{-2} + \sum_{i=1}^{N}\left[a_j^2 W_{ij}\right]\right\}^{-1}.$$

Step 4: Sample the ability parameter θ_i for each examinee i. The conditional prior distribution of θ_i is assumed to follow a normal distribution with mean $\mu_{\theta|\tau} = \mu_\theta + \sigma_{\theta\tau}\sigma_\tau^{-2}\left(\tau_i - \mu_\tau\right)$ and $\sigma_{\theta|\tau}^2 = \sigma_\theta^2 - \sigma_{\theta\tau}\sigma_\tau^{-2}\sigma_{\tau\theta}$. That is, $\theta_i \sim N\left(\mu_{\theta|\tau}, \sigma_{\theta|\tau}^2\right)$. Given $\boldsymbol{Y}$, $\boldsymbol{W}$, $\boldsymbol{a}$ and $\boldsymbol{b}$, the fully conditional posterior distribution of θ_i is given by

$$p\left(\theta_i \mid \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}\right)$$
$$\propto \prod_{j=1}^{J}\left\{\frac{\left\{\exp\left[a_j\left(\theta_i - b_j\right)\right]\right\}^{Y_{ij}}}{1 + \exp\left[a_j\left(\theta_i - b_j\right)\right]}\left[\cosh\left(2^{-1}\left|a_j\left(\theta_i - b_j\right)\right|\right)\right]\right.$$
$$\left. \times \exp\left[-\frac{\left[a_j\left(\theta_i - b_j\right)\right]^2 W_{ij}}{2}\right]\right\} p\left(\theta_i \mid \mu_{\theta|\tau}, \sigma_{\theta|\tau}^2\right).$$

Therefore, the fully conditional posterior distribution of θ_i follows a normal distribution with mean

$$\text{Var}_{\theta_i} \times \left(\mu_{\theta|\tau}\sigma_{\theta|\tau}^{-2} + \sum_{j=1}^{J}\left[a_j^2 W_{ij}\right]\left(\frac{\sum_{j=1}^{J}\left(2Y_{ij}a_j + 2a_j^2 b_j W_{ij} - a_j\right)}{2\sum_{j=1}^{J}\left[a_j^2 W_{ij}\right]}\right)\right)$$

and variance

$$\text{Var}_{\theta_i} = \left\{\sigma_{\theta|\tau}^{-2} + \sum_{j=1}^{J}\left[a_j^2 W_{ij}\right]\right\}^{-1}.$$

Step 5: Sample the response behavior variable η_{ij}. The fully conditional posterior distribution of η_{ij} is a Bernoulli distribution with success probability

$$\frac{\pi_i g_j h\left(T_{ij}; \mu_c, \sigma_c^2\right)}{\pi_i g_j h\left(T_{ij}; \mu_c, \sigma_c^2\right) + (1 - \pi_i) P\left(Y_{ij} = 1 \,|\, \theta_i, a_j, b_j\right) f\left(T_{ij}; \lambda_j, \tau_i, \sigma_j^2\right)}, \quad \text{if } Y_{ij} = 1,$$

$$\frac{\pi_i (1 - g_j) h\left(T_{ij}; \mu_c, \sigma_c^2\right)}{\pi_i (1 - g_j) h\left(T_{ij}; \mu_c, \sigma_c^2\right) + (1 - \pi_i) P\left(Y_{ij} = 0 \,|\, \theta_i, a_j, b_j\right) f\left(T_{ij}; \lambda_j, \tau_i, \sigma_j^2\right)}, \quad \text{if } Y_{ij} = 0.$$

Step 6: Sample π_i. Given a $Beta\left(\iota_1, \iota_2\right)$ prior and $\sum_{j=1}^{J} \eta_{ij} \sim$ $Binomial\left(J, \pi_i\right)$, the fully conditional posterior of π_i is

$$\pi_i \sim Beta\left(\iota_1 + \sum_{j=1}^{J} \eta_{ij}, \ \iota_2 + J - \sum_{j=1}^{J} \eta_{ij}\right).$$

Step 7: Sample g_j. Given a $Beta\left(\iota_3, \iota_4\right)$ prior, within the guessing behavior category $\eta_{ij} = 1$, the total number of people engaging in rapid guessing behavior on item j is $\sum_{i=1}^{N} \eta_{ij}$, and the number of correct items is $\sum_{i=1}^{N} \eta_{ij} Y_{ij}$; thus, $\sum_{i=1}^{N} \eta_{ij} Y_{ij} \sim$ $Binomial\left(\sum_{i=1}^{N} \eta_{ij}, g_j\right)$. The fully conditional posterior is

$$g_j \sim Beta\left(\iota_3 + \sum_{i=1}^{N} \eta_{ij} Y_{ij}, \ \iota_4 + \sum_{i=1}^{N} \eta_{ij} - \sum_{i=1}^{N} \eta_{ij} Y_{ij}\right).$$

Step 8: Sample τ_i. The conditional prior distribution of τ_i is assumed to follow a normal distribution with mean $\mu_{\tau|\theta} = \mu_\tau + \sigma_{\tau\theta} \sigma_\theta^{-2} (\theta_i - \mu_\theta)$ and $\sigma_{\tau|\theta}^2 = \sigma_\tau^2 - \sigma_{\tau\theta} \sigma_\theta^{-2} \sigma_{\theta\tau}$. That is, $\tau_i \sim N\left(\mu_{\tau|\theta}, \sigma_{\tau|\theta}^2\right)$. The fully conditional posterior distribution of τ_i given $\boldsymbol{T}^{obs}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \sigma_j^2, \boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P, \boldsymbol{\eta}$ is proportional to

$$\prod_{j=1}^{J} f\left(t_{ij}; \lambda_{jv}, \tau_i, \sigma_j^2\right)^{(1-\eta_{ij})} p\left(\tau_i \,\Big|\, \mu_{\theta|\tau}, \sigma_{\theta|\tau}^2\right).$$

The fully conditional posterior distribution of τ_i is

$$N\left(\sigma_{\tau_i^*}^2 \left(\frac{\sigma_{\theta\tau} \theta_i}{\sigma_\tau^2 - \sigma_{\theta\tau}^2} + \sum_{j=1}^{J}\left[(1 - \eta_{ij}) \sigma_j^{-2} (\lambda_j - \log t_{ij})\right]\right), \sigma_{\tau_i^*}^2\right),$$

where $\sigma_{\tau_i^*}^2 = \left((\sigma_\tau^2 - \sigma_{\theta\tau}^2)^{-1} + \sum_{j=1}^{J}\left[(1 - \eta_{ij}) \sigma_j^{-2}\right]\right)^{-1}$.

Step 9: Sample λ_j. The fully conditional posterior distribution of the intensity parameter given the parameters $\boldsymbol{T}^{obs}, \boldsymbol{\tau}, \sigma_j^2, \boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I, \boldsymbol{\eta}$ is

$$p\left(\lambda_j \,\Big|\, T_j^{obs}, \boldsymbol{\tau}, \sigma_j^2, \mu_\lambda, \sigma_\lambda^2, \boldsymbol{\eta}\right)$$
$$\propto \prod_{i=1}^{N} f\left(t_{ij}; \lambda_j, \tau_i, \sigma_j^2\right)^{(1-\eta_{ij})} p\left(\lambda_j \,|\, \mu_\lambda, \sigma_\lambda^2\right),$$

where $\lambda_j \sim N\left(\mu_\lambda, \sigma_\lambda^2\right)$. The fully conditional posterior distribution of λ_j is

$$N\left(\sigma_{\lambda_j^*}^2 \left(\mu_\lambda \sigma_\lambda^{-2} + \sum_{i=1}^{N}\left(1 - \eta_{ij}\right)\left(\log t_{ij} + \tau_i\right) \sigma_j^{-2}\right), \sigma_{\lambda_j^*}^2\right),$$

where $\sigma_{\lambda_j^*}^2 = \left(\sigma_\lambda^{-2} + \sigma_j^{-2} \sum_{i=1}^{N}\left(1 - \eta_{ij}\right)\right)^{-1}$.

Step 10: Sample σ_j^2. A prior for σ_j^2 is an inverse-gamma distribution, $IG\left(\upsilon_1, \omega_1\right)$. The fully conditional posterior distribution of σ_j^2 is

$$IG\left(\upsilon_1 + \frac{\sum_{i=1}^{N}\left(1 - \eta_{ij}\right)}{2}, \omega_1 + \frac{\sum_{i=1}^{N}\left[\left(1 - \eta_{ij}\right)\left(\log t_{ij} - \lambda_j + \tau_i\right)^2\right]}{2}\right).$$

Step 11: Sample μ_c. We assume a uniform prior $p\left(\mu_c\right) \propto 1$. The fully conditional posterior distribution of μ_c is proportional to

$$p\left(\mu_c \,|\, \boldsymbol{T}, \boldsymbol{\eta}\right) \propto \prod_{i=1}^{N} \prod_{j=1}^{J} f\left(t_{ij}; \mu_c, \sigma_c^2\right)^{\eta_{ij}} p\left(\mu_c\right).$$

The fully conditional posterior distribution of μ_c is

$$\mu_c \,|\, \boldsymbol{T}, \boldsymbol{\eta} \sim N\left(\left(\sum_{i=1}^{N} \sum_{j=1}^{J} \eta_{ij}\right)^{-1} \left(\sum_{i=1}^{N} \sum_{j=1}^{J} \eta_{ij} \log t_{ij}\right),\right.$$
$$\left.\left(\sum_{i=1}^{N} \sum_{j=1}^{J} \eta_{ij}\right)^{-1} \sigma_c^2\right).$$

Step 12: Sample σ_c^2. We assume that the variance parameter follows an inverse-gamma prior distribution, $IG\left(\upsilon_2, \omega_2\right)$. The fully conditional posterior distribution of σ_c^2 given $\boldsymbol{T}$, $\mu_c, \upsilon_2, \omega_2, \boldsymbol{\eta}$ is proportional to

$$p\left(\sigma_c^2 \,|\, \boldsymbol{T}, \mu_c, \upsilon_2, \omega_2, \boldsymbol{\eta}\right) \propto \prod_{i=1}^{N} \prod_{j=1}^{J} f\left(t_{ij}; \mu_c, \sigma_c^2\right)^{\eta_{ij}} p\left(\sigma_c^2\right).$$

The fully conditional posterior distribution of σ_c^2 is

$$\sigma_c^2 \,|\, \boldsymbol{T}, \mu_c, \upsilon_2, \omega_2, \boldsymbol{\eta}$$
$$\sim IG\left(\upsilon_1 + \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \eta_{ij}}{2}, \omega_1 + \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \eta_{ij}\left(\log t_{ij} - \mu_c\right)^2}{2}\right).$$

3.2. Metropolis–Hastings Sampling Algorithm

In order to estimate the constrained covariance matrix $\boldsymbol{\Sigma}_P = \begin{pmatrix} 1 & \sigma_{\theta\tau} \\ \sigma_{\tau\theta} & \sigma_\tau^2 \end{pmatrix}$ (where σ_θ^2 is restricted to be equal to 1 owing to the model identification limitation), we need to update each element of the constrained covariance matrix separately using the Metropolis–Hastings algorithm.

Step 13: Sample the correlation $\sigma_{\theta\tau}$ between $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. The identification constraints induce a restricted covariance matrix. The new value $\sigma_{\theta\tau}^*$ is sampled from a truncated normal distribution $N\left(\sigma_{\theta\tau}^{(r-1)}, s_{01}^2\right) I\left(-p_{01} < \sigma_{\theta\tau}^* < p_{01}\right)$, where $p_{01} = \sqrt{\sigma_\tau^{2,(r-1)}}$. Therefore, the probability of acceptance $\alpha\left(\sigma_{\theta\tau}^{(r-1)}, \sigma_{\theta\tau}^*\right)$ can be written as

$$\min\left\{1, \frac{\prod_{i=1}^N p\left(\tau_i \Big| \theta_i^{(r)}, \sigma_\tau^{2,(r-1)}, \sigma_{\theta\tau}^*\right) p\left(\sigma_{\theta\tau}^*\right) \left(\Phi\left(\frac{p_{01}-\sigma_{\theta\tau}^{(r-1)}}{s_{01}}\right) - \Phi\left(\frac{-p_{01}-\sigma_{\theta\tau}^{(r-1)}}{s_{01}}\right)\right)}{\prod_{i=1}^N p\left(\tau_i \Big| \theta_i^{(r)}, \sigma_\tau^{2,(r-1)}, \sigma_{\theta\tau}^{(r-1)}\right) p\left(\sigma_{\theta\tau}^{(r-1)}\right) \left(\Phi\left(\frac{p_{01}-\sigma_{\theta\tau}^*}{s_{01}}\right) - \Phi\left(\frac{-p_{01}-\sigma_{\theta\tau}^*}{s_{01}}\right)\right)}\right\};$$

otherwise, $\sigma_{\theta\tau}^{(r-1)} = \sigma_{\theta\tau}^*$, where $p\left(\tau_i | \theta_i\right)$ is the conditional density function of the speed parameter, s_{01}^2 is the proposal variance, and $p\left(\sigma_{\theta\tau}\right)$ is the density of the uniform prior.

Step 14: Sample σ_τ^2. The new value $\sigma_\tau^{2,*}$ is sampled from a truncated normal distribution $N\left(\sigma_\tau^{2,(r-1)}, s_{02}^2\right) I\left(\sigma_\tau^{2,*} > \left(\sigma_{\theta^{(2)}\tau}^*\right)^2 = p_0\right)$. Therefore, the probability of acceptance $\alpha\left(\sigma_\tau^{2,(r-1)}, \sigma_\tau^{2,*}\right)$ can be written as

$$\min\left\{1, \frac{\prod_{i=1}^N p\left(\tau_i \Big| \theta_i^{(r)}, \sigma_\tau^{2,*}, \sigma_{\theta\tau}^{(r)}\right) p\left(\sigma_\tau^{2,*}; \kappa, \vartheta\right) \left(1 - \Phi\left(\frac{p_0-\sigma_\tau^{2,(r-1)}}{s_{02}}\right)\right)}{\prod_{i=1}^N p\left(\tau_i \Big| \theta_i^{(r)}, \sigma_\tau^{2,(r-1)}, \sigma_{\theta\tau}^{(r)}\right) p\left(\sigma_\tau^{2,(r-1)}; \kappa, \vartheta\right) \left(1 - \Phi\left(\frac{p_0-\sigma_\tau^{2,*}}{s_{02}}\right)\right)}\right\};$$

otherwise, $\sigma_\tau^{2,*} = \sigma_\tau^{2,(r-1)}$, where s_{02}^2 is the proposal variance, and $p\left(\sigma_\tau^2; \kappa, \vartheta\right)$ is the density function of the scaled inverse chi-squared distribution with degrees of freedom and the scale parameter.

3.3. Bayesian Model Assessment

Two Bayesian model assessment methods were developed to evaluate the fit of the two models. The new model is a mixture model that combines responses and response times to detect rapid guessing behavior. The other model does not consider the mixture structure. Spiegelhalter et al. (2002) proposed the deviance information criterion (DIC) as a way to evaluate model fit based on Bayesian posterior estimates by considering the trade-off relationship between the adequacy of the model fitting and the number of model parameters. Write $\boldsymbol{\Lambda} = \left(\boldsymbol{\Lambda}_{ij}, i = 1, ..., N. j = 1, ..., J.\right)$, where $\boldsymbol{\Lambda}_{ij} = \left(\eta_{ij}, \theta_i, a_j, b_j, \lambda_j, \tau_i, \sigma_j^2, \mu_c, \sigma_c^2, g_j, \pi_i\right)'$. Let $\left\{\boldsymbol{\Lambda}^{(1)}, ..., \boldsymbol{\Lambda}^{(M)}\right\}$, where $\boldsymbol{\Lambda}^{(m)} =$

$\left(\eta_{ij}^{(m)}, \theta_i^{(m)}, a_j^{(m)}, b_j^{(m)}, \lambda_j^{(m)}, \tau_i^{(m)}, \sigma_j^{2,(m)}, \mu_c^{(m)}, \sigma_c^{2,(m)}, g_j^{(m)}, \pi_i^{(m)}\right)'$ for $m = 1, ..., M$, denotes an Markov chain Monte Carlo (MCMC) sample from the posterior distribution in (1). The logarithm of the joint likelihood function evaluated at $\boldsymbol{\Lambda}^{(m)}$ is given by

$$\log f\left(\boldsymbol{Y}, \boldsymbol{T} \Big| \boldsymbol{\Lambda}^{(m)}\right) = \sum_{i=1}^N \sum_{j=1}^J \log f\left(Y_{ij}, T_{ij} \Big| \boldsymbol{\Lambda}_{ij}^{(m)}\right), \quad (4)$$

where

$$f\left(Y_{ij}, T_{ij} \Big| \boldsymbol{\Lambda}_{ij}\right)$$
$$= \left[\pi_i g_j h\left(T_{ij} | \mu_c, \sigma_c^2\right)\right]^{\eta_{ij}\cdot Y_{ij}} \left[\pi_i \left(1 - g_j\right) h\left(T_{ij} | \mu_c, \sigma_c^2\right)\right]^{\eta_{ij}\cdot(1-Y_{ij})}$$

$$\times \left[\left(1 - \pi_i\right) P\left(Y_{ij} = 1 \Big| \eta_{ij} = 0\right) f\left(T_{ij} \Big| \lambda_j, \tau_i, \sigma_j^2\right)\right]^{(1-\eta_{ij})\cdot Y_{ij}}$$
$$\times \left[\left(1 - \pi_i\right) P\left(Y_{ij} = 0 \Big| \eta_{ij} = 0\right) f\left(T_{ij} \Big| \lambda_j, \tau_i, \sigma_j^2\right)\right]^{(1-\eta_{ij})\cdot(1-Y_{ij})}.$$

As the log-likelihood function $\log f\left(Y_{ij}, T_{ij} \Big| \boldsymbol{\Lambda}_{ij}^{(r)}\right)$, $i = 1, ..., N. j = 1, ..., J$, is readily available from the R outputs, $\log f\left(\boldsymbol{Y}, \boldsymbol{T} | \boldsymbol{\Lambda}^{(r)}\right)$ in (4) is easy to compute. The DIC can be calculated as follows:

$$\text{DIC} = \widehat{\text{Dev}(\boldsymbol{\Lambda})} + 2P_D = \widehat{\text{Dev}(\boldsymbol{\Lambda})} + 2\left[\overline{\text{Dev}(\boldsymbol{\Lambda})} - \widehat{\text{Dev}(\boldsymbol{\Lambda})}\right], \quad (5)$$

where

$$\overline{\text{Dev}(\boldsymbol{\Lambda})} = -\frac{2}{M} \sum_{m=1}^M \log f\left(\boldsymbol{Y}, \boldsymbol{T} \Big| \boldsymbol{\Lambda}^{(m)}\right) \text{ and}$$

$$\widehat{\text{Dev}(\boldsymbol{\Lambda})} = -2 \max_{1 \leq m \leq M} \log f\left(\boldsymbol{Y}, \boldsymbol{T} \Big| \boldsymbol{\Lambda}^{(m)}\right).$$

In (5), $\overline{\text{Dev}(\boldsymbol{\Lambda})}$ is a Monte Carlo estimate of the posterior expectation of the deviance function $\text{Dev}(\boldsymbol{\Lambda}) = -2 \log f\left(\boldsymbol{Y}, \boldsymbol{T} | \boldsymbol{\Lambda}\right)$. $\widehat{\text{Dev}(\boldsymbol{\Lambda})}$ is an approximation of $\text{Dev}(\hat{\boldsymbol{\Lambda}})$, where $\hat{\boldsymbol{\Lambda}}$ is the posterior mode, when the prior is relatively non-informative, and $P_D = \overline{\text{Dev}(\boldsymbol{\Lambda})} - \widehat{\text{Dev}(\boldsymbol{\Lambda})}$ is the effective number of parameters. The model with a smaller DIC value fits the data better.

Another method to compare the fit of the two models is to use the logarithm of the pseudomarginal likelihood (LPML; Geisser and Eddy, 1979; Ibrahim et al., 2001) by calculating the conditional predictive ordinates (CPO) index. Next, the formulas for computing the CPO and LPML are given. Letting $U_{ij,\max} = \max_{1 \le m \le M} \left\{ -\log f \left(Y_{ij}, T_{ij} \Big| \mathbf{\Lambda}_{ij}^{(m)} \right) \right\}$, a Monte Carlo estimate of the CPO (Gelfand et al., 1992; Chen et al., 2000) is given by

$$
\begin{aligned}
\log \left(\widehat{\text{CPO}_{ij}} \right) = & -U_{ij,\max} \\
& - \log \left[\frac{1}{M} \sum_{m=1}^{M} \exp \left\{ -\log f \left(Y_{ij}, T_{ij} \Big| \mathbf{\Lambda}_{ij}^{(m)} \right) - U_{ij,\max} \right\} \right] .(6)
\end{aligned}
$$

Note that the maximum value adjustment used in $\log \left(\widehat{\text{CPO}_{ij}} \right)$ plays an important part in numerical stabilization when computing $\exp \left\{ -\log f \left(Y_{ij}, T_{ij} \Big| \mathbf{\Lambda}_{ij}^{(m)} \right) - U_{ij,\max} \right\}$ in (6). A summary statistic of the $\widehat{\text{CPO}_{ij}}$ is the sum of their logarithms, which is called the LPML and given by

$$
\text{LPML} = \sum_{i=1}^{N} \sum_{j=1}^{J} \log \left(\widehat{\text{CPO}_{ij}} \right).
$$

A model with a larger LPML has a better fit to the data.

4. SIMULATION STUDIES

4.1. Simulation 1

This simulation study was conducted to evaluate the recovery performance of the Pólya–gamma Gibbs sampling algorithm under different simulation conditions.

Simulation Designs

The following conditions were manipulated: (a) test length, $J = 20$ or 40, where the 20-item test is within 40 min, and the 40-item test is within 80 min; (b) the number of examinees, $N = 1,000$ or $2,000$; and (c) the speededness level, low speededness level (LSL) or high speededness level (HSL). The speededness level is controlled by the intensity parameter λ_j. That is, a larger time intensity parameter corresponds to a longer average testing time. Fully crossing the different values of these four factors yielded eight conditions (two test lengths × two sample sizes × two speededness levels).
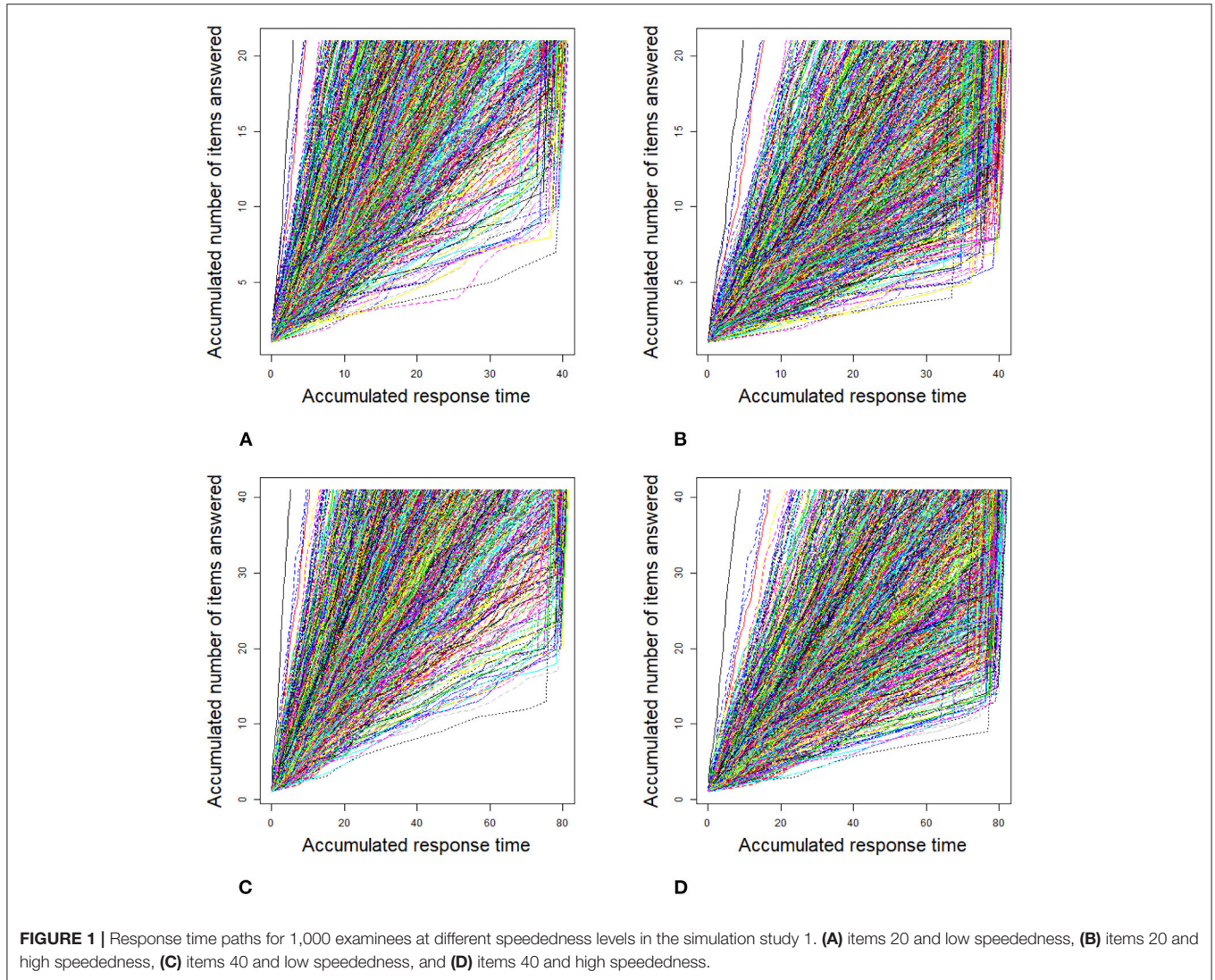
True Values and Prior Distributions

For the 2PL model, true values of item discrimination parameters a_j are generated from a truncated normal distribution, i.e., $a_j \sim N(0,1) \, \text{I}(0, +\infty)$, $j = 1, 2, ..., J$, where the indicator function $\text{I}(A)$ takes a value of 1 if A is true and 0 if A is false. The item difficulty parameters b_j are generated from a standardized normal distribution. For the RT model, the response times of the rapid guessing behavior, C_{ij}, are generated from a log-normal distribution (Wang and Xu, 2015, p. 464), i.e., $\log C_{ij} \sim N(-2, 0.25)$. The correct response probability of the rapid guessing behavior, g_j, is set to 0.25 for all items (Wang and Xu, 2015). Although the variances of the RT model, σ_j^2, can vary across items in the process of model setting and algorithm

TABLE 1 | The proportions of examinees and items in the simulation study 1.

	No. of items = 20	
	No. of examinees 1,000	No. of examinees 2,000
Item intensity	Proportion of examinees who can not finish a 20 item test within 40 min	
$\lambda \sim U(-0.25, 0.25)$	14.2%	12.3%
$\lambda \sim U(0.25, 0.75)$	46.6%	40.5%
Item intensity	Proportion of items that are answered by rapid guessing	
$\lambda \sim U(-0.25, 0.25)$	3.31%	2.88%
$\lambda \sim U(0.25, 0.75)$	14.86%	12.59%
	No. of items = 40	
	No. of examinees 1,000	No. of examinees 2,000
Item intensity	Proportion of examinees who can not finish a 40 item test within 80 min	
$\lambda \sim U(-0.25, 0.25)$	13.4%	15.4%
$\lambda \sim U(0.25, 0.75)$	44.1%	47.9%
Item intensity	Proportion of items that are answered by rapid guessing	
$\lambda \sim U(-0.25, 0.25)$	3.05%	3.43%
$\lambda \sim U(0.25, 0.75)$	13.90%	14.61%

implementation, for convenience, we assume that the variance of the RT model, σ_j^2, is set to 0.5 for all items. We controlled the speededness level by adjusting the time intensity parameter, that is, low speededness distribution $\lambda \sim U(-0.25, 0.25)$ and high speededness distribution $\lambda \sim U(0.25, 0.75)$. The proportion of examinees who could not finish a test within the allocated time is shown in **Table 1**. The proportion of items that were answered by guessing is also shown in **Table 1**. For the population distribution of person parameters, the ability and speed parameters $(\theta, \tau)'$ were generated from a bivariate normal distribution with mean vector $(0, 0)'$ and covariance matrix $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.25 \end{pmatrix}$. The responses and response times were generated from the 2PL model and log-normal distribution. The following method was used to generate the guessing behavior indicator η_{ij}. For all items, examinees could finish a given test within the allotted time having $\eta_{ij} = 0$, where $j = 1, ..., J$. Other η_{ij} were generated by the following two steps. Assuming that the generated response time data has no time limit for all items, then we replace T_{ij} with C_{ij} from the last item backward until the total response time is less than or equal to the allocated time. Therefore, given the eight simulation conditions, the RT paths for the examinees are shown in **Figures 1, 2**. **Figures 3, 4** show the histograms of response times obtained from all item–person combinations. The non-informative priors and hyperpriors for the parameters were chosen as follows: $p(a_j) \sim N(0, 10^5) \, \text{I}(0, +\infty)$,

FIGURE 1 | Response time paths for 1,000 examinees at different speededness levels in the simulation study 1. **(A)** items 20 and low speededness, **(B)** items 20 and high speededness, **(C)** items 40 and low speededness, and **(D)** items 40 and high speededness.

$p(b_j) \sim N(0, 10^5)$, $p(g_j) \sim Beta(5, 17)$, $p(\lambda_j) \sim N(0, 10^5)$, $p(\pi_i) \sim Beta(1, 5)$, $p(\mu_c) \sim N(-3, 10^5)$, $p(\sigma_c^2) \sim Inv\text{-}Gamma(0.0001, 0.0001)$, $p(\sigma_\tau^2) \sim Inv\text{-}Gamma(0.0001, 0.0001)$, and $\sigma_{\theta\tau} \sim U\left(-\sqrt{\sigma_\tau^2}, \sqrt{\sigma_\tau^2}\right)$, where $\sigma_\tau^2 = 1$. Fifty replications were considered in each simulation condition.

Convergence diagnostics

In order to evaluate the convergence of parameter estimates, we only considered convergence in the case of minimum sample sizes with HSLs owing to space limitations. That is, the test length was fixed at 20, and the number of examinees was 1,000. Two methods were used to check the convergence of our algorithm: the "eyeball" method to monitor the convergence by visually inspecting the history plots of the generated sequences; and the Gelman–Rubin method (Gelman and Rubin, 1992; Brooks and Gelman, 1998).

The convergence of the Bayesian algorithm was checked by monitoring the trace plots of the parameters for consecutive sequences of 20,000 iterations. The first 10,000 iterations were set as the burn-in period. As an illustration, four chains started at overdispersed starting values were run for each replication. The trace plots of item parameters randomly selected are shown in **Figure 5**. In addition, the potential scale reduction factor (PSRF; Brooks and Gelman, 1998) values for all item parameters are shown in **Figure 6**. We found that the PSRF values for all parameters were less than 1.2, which ensured that all chains converged as expected.

5. RESULTS

As shown in **Table 2**, the bias was between 0.0098 and 0.1411 for the discrimination parameters $\boldsymbol{a}$, between −0.0335 and 0.0010 for the difficulty parameters $\boldsymbol{b}$, between −0.0206 and 0.0115 for the rapid guessing parameters $\boldsymbol{g}$, between −0.0271 and 0.0386 for the time intensity parameters $\boldsymbol{\lambda}$, between −0.0105 and 0.0314 for the time discrimination parameters σ^2, between 0.0196 and 0.0313 for the ability parameters $\boldsymbol{\theta}$, between 0.0058 and 0.0377 for

FIGURE 2 | Response time paths for 2,000 examinees at different speededness levels in the simulation study 1. **(A)** items 20 and low speededness, **(B)** items 20 and high speededness, **(C)** items 40 and low speededness, and **(D)** items 40 and high speededness.
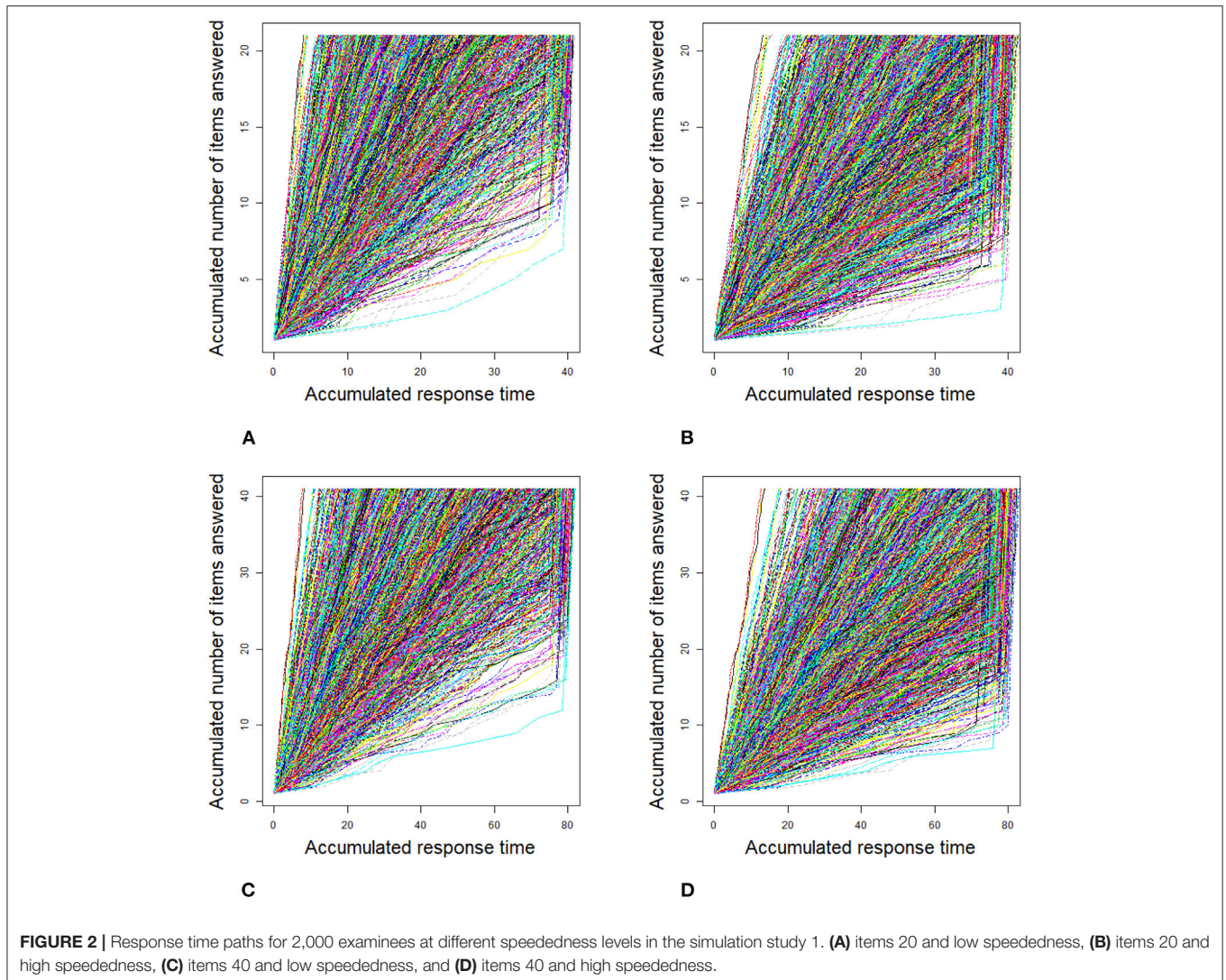
the speed parameters τ, between -0.0259 and 0.0202 for the μ_c, between -0.0373 and 0.0136 for the σ_c^2, between -0.0671 and -0.0102 for the $\sigma_{\theta\tau}$, and between -0.0201 and 0.0056 for the σ_τ^2. In addition, the MSE was between 0.0125 and 0.0413 for the discrimination parameters $\boldsymbol{a}$, between 0.0041 and 0.0138 for the difficulty parameters $\boldsymbol{b}$, between 0.0009 and 0.0026 for the rapid guessing parameters $\boldsymbol{g}$, between 0.0001 and 0.0017 for the time intensity parameters λ, between 0.0001 and 0.0005 for the time discrimination parameters $\boldsymbol{\sigma}^2$, between 0.0873 and 0.1920 for the ability parameters $\boldsymbol{\theta}$, between 0.0068 and 0.0693 for the speed parameters $\boldsymbol{\tau}$, between 0.0000 and 0.0007 for the μ_c, between 0.0000 and 0.0009 for the σ_c^2, between 0.0010 and 0.0045 for the $\sigma_{\theta\tau}$, and between 0.0002 and 0.0009 for the σ_τ^2. In summary, the Pólya–gamma Gibbs sampling algorithm provides accurate estimates of the parameters for various numbers of examinees and items.

5.1. Simulation 2

In this simulation study, we focus on the model fitting data for the mixture model and non-mixture model based on different simulation conditions from the perspective of Bayesian model assessment. Two Bayesian model assessment tools, DIC and LPML, are used to identify the true models.

Simulation Designs

For purposes of illustration, the numbers of examinees and items were fixed at 1,000 and 40, respectively. The true value settings for the item parameters in the 2PLIRT model and response time model were the same as in simulation study 1. The first factor is the correlation coefficient. Three correction coefficients $\rho_{\theta\tau}$ were considered in this simulation. That is, (1) $\rho_{\theta\tau} = 0.3$ (θ and τ have weak correlation; WC); (2) $\rho_{\theta\tau} = 0.8$ (θ and τ have a strong correlation; SC). Furthermore, the true values of θ and τ can be drawn from a bivariate
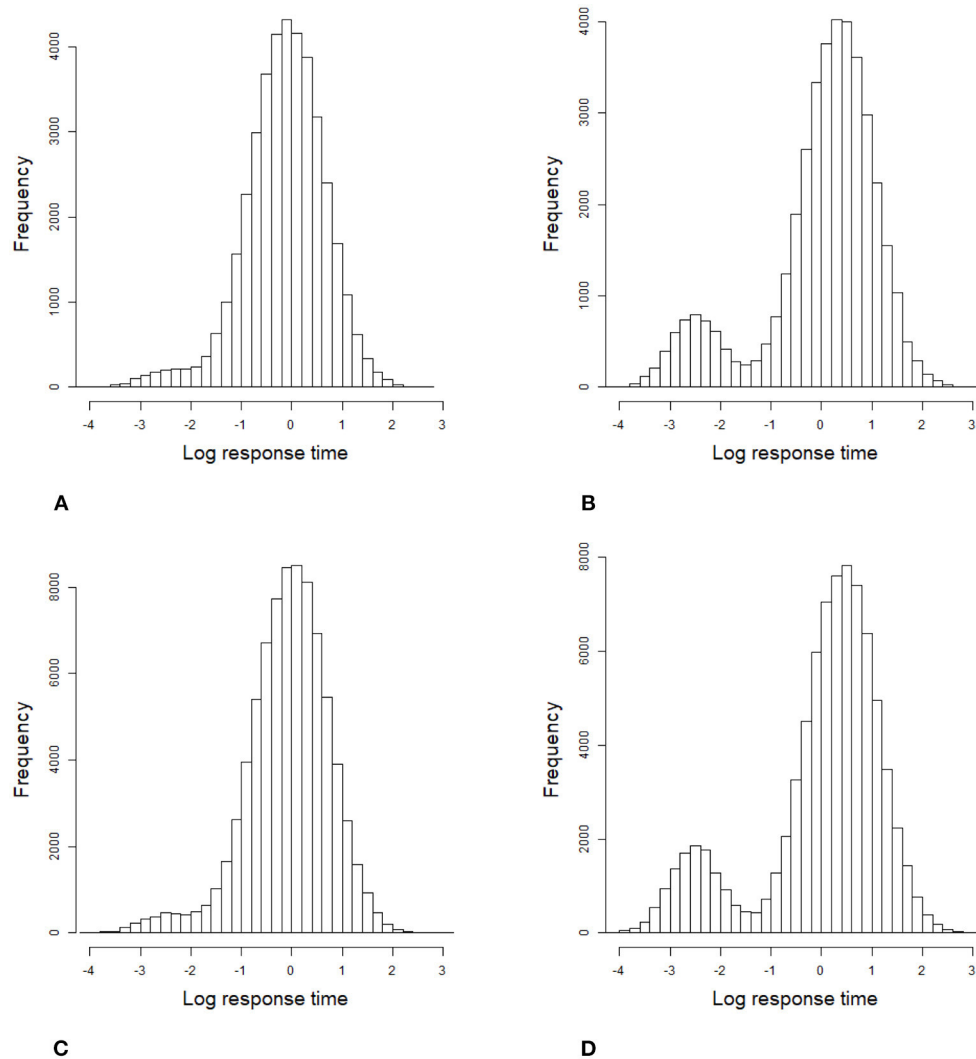
FIGURE 3 | Histogram of 1,000 examinees' response times based on all item–person combinations in the simulation study 1. **(A)** items 20 and low speededness, **(B)** items 20 and high speededness, **(C)** items 40 and low speededness, and **(D)** items 40 and high speededness.
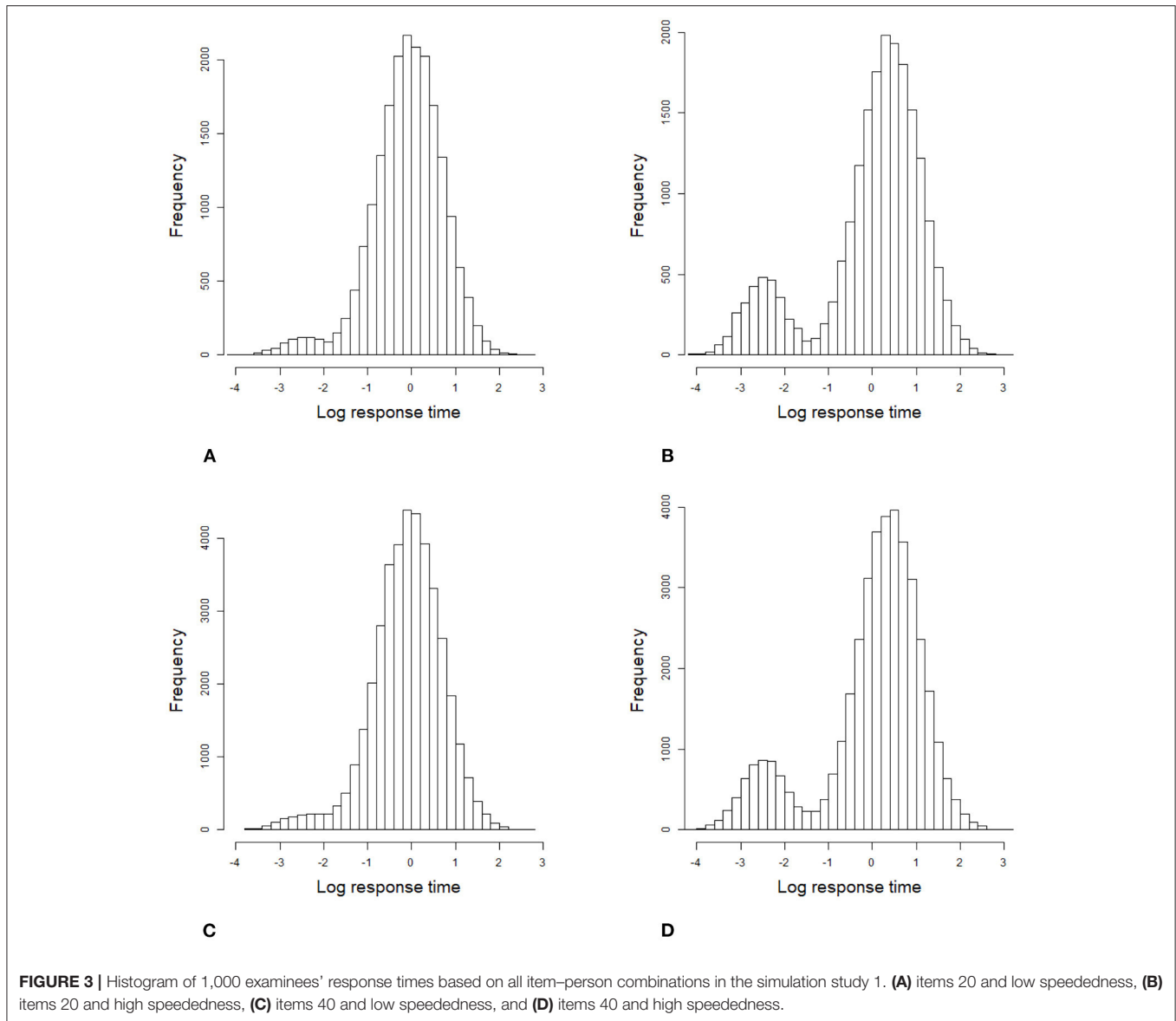
normal distribution with mean vector **0** and covariance matrix $\begin{pmatrix} 1 & \rho_{\theta\tau} \\ \rho_{\theta\tau} & 1 \end{pmatrix}$. The second factor is the speededness level, which was varied by adjusting the time intensity parameter λ: (1) LSL, $\lambda \sim U(-0.25, 0.25)$; (2) HSL, $\lambda \sim U(0.25, 0.75)$. The third factor is the choice of fitting model: (1) mixture model; (2) non-mixture model (hierarchical structure model of van der Linden, 2007). Based on the abovementioned test conditions, the item responses and response time data were respectively generated from the 2PLIRT model and response time model. Therefore, the true models and the fitted models were designed as follows.

(i) True model, i.e., mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕LSL **vs.** fitted model, i.e., mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕LSL, and non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕LSL.

(ii) True model, i.e., mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕LSL **vs.** fitted model, i.e., mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕LSL, and non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕LSL.

(iii) True model, i.e., mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕HSL **vs.** fitted model, i.e., mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕HSL, and non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕HSL.

(iv) True model, i.e., mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕HSL **vs.** fitted model, i.e., mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕HSL, and non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕HSL.

(v) True model, i.e., non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕LSL **vs.** fitted model, i.e., mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕LSL, and non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕LSL.

FIGURE 4 | Histogram of 2,000 examinees' response times based on all item–person combinations in the simulation study 1. **(A)** items 20 and low speededness, **(B)** items 20 and high speededness, **(C)** items 40 and low speededness, and **(D)** items 40 and high speededness.

(vi) True model, i.e., non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕LSL **vs.** fitted model, i.e., mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕LSL, and non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕LSL.

(vii) True model, i.e., non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕HSL **vs.** fitted model, i.e., mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕HSL, and non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)⊕HSL.

(viii) True model, i.e., non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕HSL **vs.** fitted model, i.e., mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕HSL, and non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)⊕HSL.

The priors of parameters were also the same as those used in simulation 1. That is, the non-informative priors were used in this simulation study. To implement the MCMC sampling algorithm, chains of length 10,000 with an initial burn-in period

of 20,000 were chosen. There were 50 replications for each simulation condition. The PSRF (Brooks and Gelman, 1998) values for all item and person parameters for each simulation condition were less than 1.2.

Results

As shown in **Tables 3**, **4**, regardless of whether the speededness levels were low or high, and whether the correlation coefficients were weak ($\rho_{\theta\tau} = 0.3$) or strong ($\rho_{\theta\tau} = 0.8$), both Bayesian model assessment criteria could accurately identify the true models when the data were generated from the mixture models and non-mixture models. More specifically, under the LSL and WC conditions, when the mixture model was the true model, the mixture model fitted the data better, as expected. The median DIC of the mixture model (185007.092) was smaller than that of the non-mixture model (201335.596), and the median LPML of the mixture model (−91302.451) was larger than that of
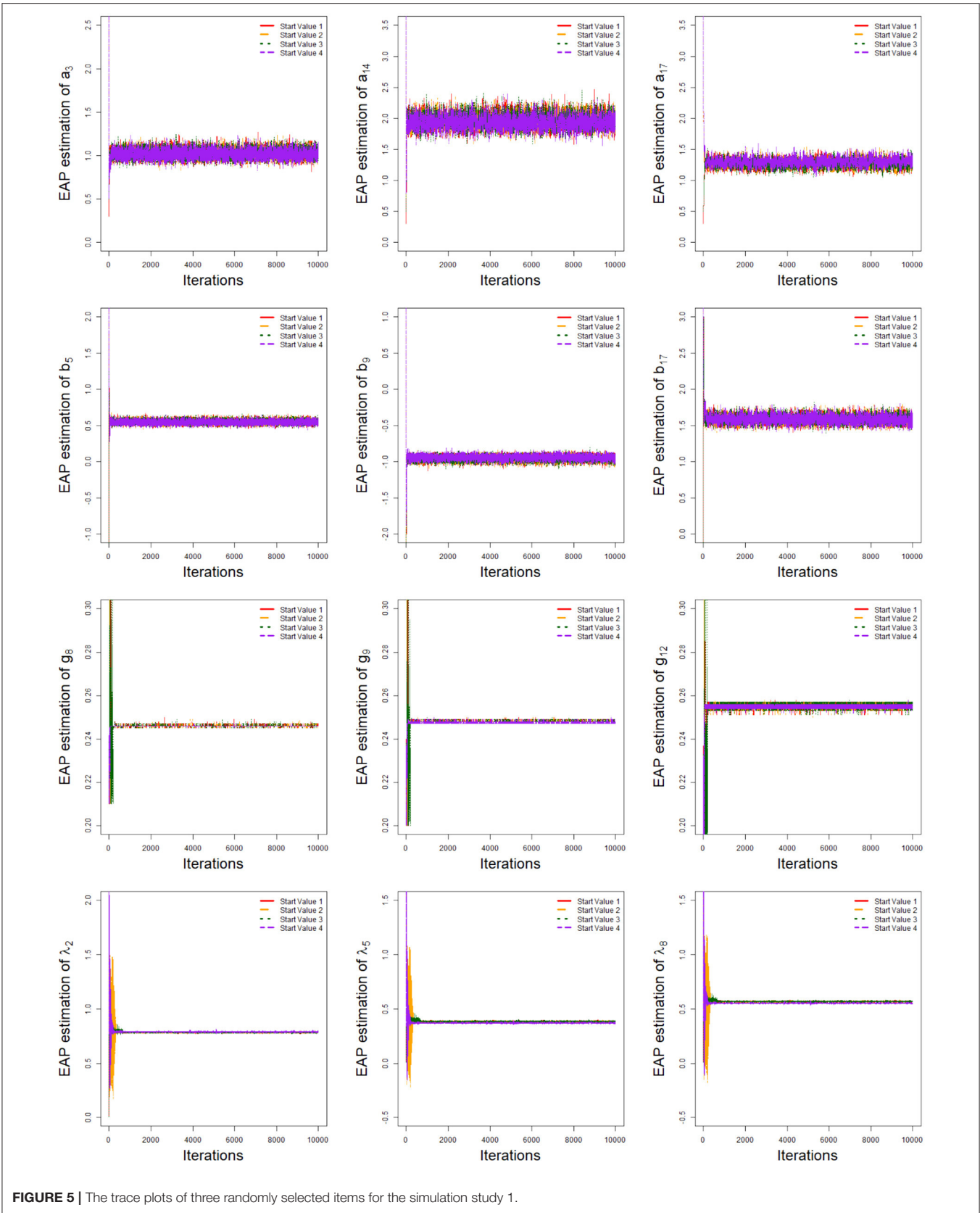
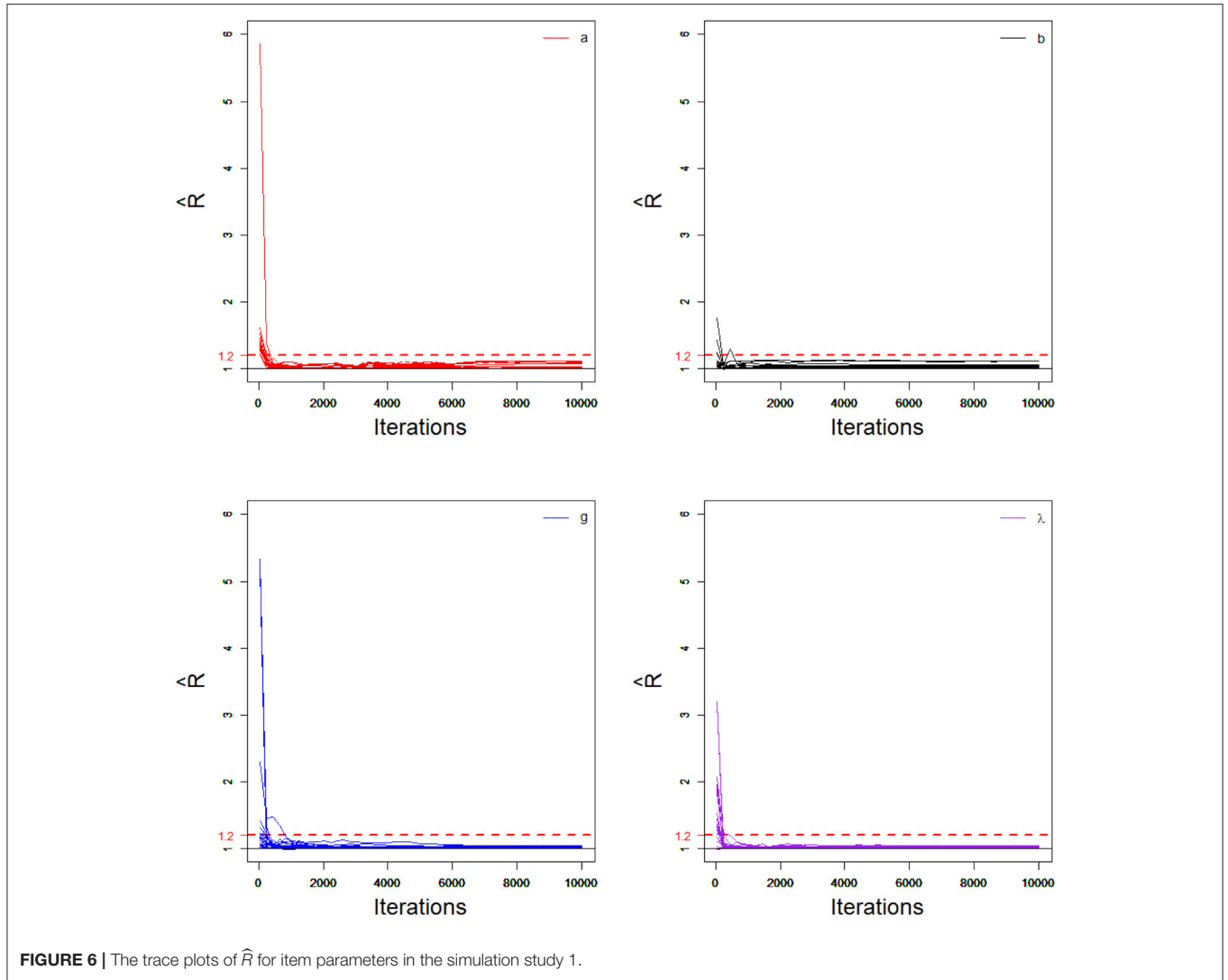**FIGURE 5 |** The trace plots of three randomly selected items for the simulation study 1.

FIGURE 6 | The trace plots of $\widehat{R}$ for item parameters in the simulation study 1.

TABLE 2 | Evaluating the accuracy of parameters based on mixture model in simulation study 1.

	$N = 1,000, J = 20$		$N = 1,000, J = 40$		$N = 2,000, J = 20$		$N = 2,000, J = 40$	
	LSL	**HSL**	**LSL**	**HSL**	**LSL**	**HSL**	**LSL**	**HSL**
Bias								
a	0.0320	0.0814	0.0098	0.0291	0.1002	0.1411	0.0253	0.0472
b	−0.0149	−0.0162	−0.0252	−0.0335	−0.0203	−0.0194	0.0010	−0.0030
g	−0.0136	−0.0203	−0.0193	−0.0166	−0.0005	−0.0022	−0.0206	0.0115
λ	0.0195	−0.0169	0.0386	0.0160	0.0077	−0.0271	0.0152	−0.0100
σ^2	−0.0105	0.0058	−0.0062	−0.0041	−0.0092	0.0123	−0.0080	0.0314
θ	0.0268	0.0295	0.0210	0.0220	0.0286	0.0313	0.0196	0.0260
τ	0.0214	0.0137	0.0377	0.0218	0.0098	0.0058	0.0168	0.0152
μ_C	−0.0259	0.0092	−0.0108	0.0078	−0.0226	0.0069	0.0041	0.0202
σ_C^2	−0.0373	0.0132	−0.0371	0.0115	−0.0371	0.0063	−0.0331	0.0136
$\sigma_{\theta\tau}$	−0.0474	−0.0671	−0.0373	−0.0501	−0.0324	−0.0552	−0.0119	−0.0102
σ_τ^2	−0.0182	−0.0201	−0.0049	−0.0103	−0.0057	−0.0054	0.0037	0.0056

(Continued)

TABLE 2 | Continued

	$N = 1,000, J = 20$		$N = 1,000, J = 40$		$N = 2,000, J = 20$		$N = 2,000, J = 40$	
	LSL	HSL	LSL	HSL	LSL	HSL	LSL	HSL
MSE								
a	0.0252	0.0355	0.0287	0.0375	0.0413	0.0527	0.0125	0.0167
b	0.0071	0.0085	0.0105	0.0138	0.0084	0.0108	0.0041	0.0058
g	0.0026	0.0018	0.0014	0.0011	0.0013	0.0010	0.0015	0.0009
λ	0.0007	0.0011	0.0017	0.0006	0.0001	0.0013	0.0003	0.0015
σ^2	0.0001	0.0004	0.0001	0.0002	0.0001	0.0005	0.0001	0.0003
θ	0.1587	0.1841	0.0943	0.1107	0.1711	0.1920	0.0873	0.1155
τ	0.0133	0.0557	0.0080	0.0141	0.0148	0.0693	0.0068	0.0099
μ_C	0.0006	0.0000	0.0001	0.0000	0.0005	0.0000	0.0000	0.0007
σ_C^2	0.0003	0.0001	0.0006	0.0001	0.0005	0.0000	0.0009	0.0001
$\sigma_{\theta\tau}$	0.0022	0.0045	0.0014	0.0025	0.0010	0.0030	0.0015	0.0026
σ_τ^2	0.0003	0.0004	0.0002	0.0005	0.0002	0.0006	0.0005	0.0009

*Note that the Bias and MSE denote the average Bias and MSE for the interested parameters. **a** represents all discrimination parameters, **b** represents all difficulty parameters, **g** represents all rapid guessing parameters, **λ** represents all time intensity parameters, **σ²** represents all time discrimination parameters, **θ** represents all ability parameters, and **τ** represents all speed parameters.*

TABLE 3 | The results of Bayesian model assessment in simulation study 2.

			Low speededness level (LSL)		
	Fitted model			**Mixture model with WC**	**Non-mixture model with WC**
	Mixture model	DIC	Q_1	183970.082	200906.367
	with WC		Median	185007.092	201335.596
True	$(\rho_{\theta\tau} = 0.3)$		Q_3	185472.819	201700.856
model			Q_1	−91433.366	−103949.160
		LPML	Median	−91302.451	−103871.796
			Q_3	−91095.166	−103782.198

			Low speededness level (LSL)		
	Fitted model			**Mixture model with SC**	**Non-mixture model with SC**
	Mixture model	DIC	Q_1	182423.016	200490.494
	with SC		Median	182806.907	200960.661
True	$(\rho_{\theta\tau} = 0.8)$		Q_3	183285.554	201204.742
Model			Q_1	−91270.116	−103687.867
		LPML	Median	−91213.797	−103584.228
			Q_3	−91100.563	−103419.208

			High speededness level (HSL)		
	Fitted model			**Mixture model with WC**	**Non-mixture model with WC**
	Mixture model	DIC	Q_1	159487.663	175985.981
	with WC		Median	159985.584	176499.862
True	$(\rho_{\theta\tau} = 0.3)$		Q_3	161227.782	176989.732
Model			Q_1	−80685.663	−87906.257
		LPML	Median	−80474.893	−87782.508
			Q_3	−80332.172	−87673.533

			High speededness level (HSL)		
	Fitted model			**Mixture model with SC**	**Non-mixture model with SC**
	Mixture model	DIC	Q_1	159235.762	175815.800
	with SC		Median	159629.846	176335.113
True	$(\rho_{\theta\tau} = 0.8)$		Q_3	160570.239	176859.457
Model			Q_1	−80840.626	−87917.891
		LPML	Median	−80736.678	−87714.244
			Q_3	−80570.342	−87638.130

Note that the mixture model is the model in Section 2. The non-mixture model is the hierarchical structure model in van der Linden (2007).

TABLE 4 | The results of Bayesian model assessment in simulation study 2.

				Low speededness level (LSL)	
	Fitted model			**Mixture model with WC**	**Non-mixture model with WC**
	Non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)	DIC	Q_1	191642.341	187822.030
			Median	192051.824	188057.725
True Model			Q_3	192465.323	188289.444
		LPML	Q_1	−95287.618	−93306.447
			Median	−95204.235	−93222.498
			Q_3	−95146.751	−93168.033

				Low speededness level (LSL)	
	Fitted model			**Mixture model with SC**	**Non-mixture model with SC**
	Non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)	DIC	Q_1	191663.580	187582.329
			Median	192059.746	187868.073
True Model			Q_3	192341.397	187988.874
		LPML	Q_1	−95293.492	−93319.285
			Median	−95177.928	-93224.461
			Q_3	−95127.793	−93132.479

				High speededness level (HSL)	
	model	Fitted		**Mixture model with WC**	**Non-mixture model with WC**
	Non-mixture model with WC ($\rho_{\theta\tau} = 0.3$)	DIC	Q_1	191880.178	187523.642
			Median	192161.323	187831.945
True Model			Q_3	192528.860	188102.832
		LPML	Q_1	−95194.438	−93202.085
			Median	−95108.402	-93129.144
			Q_3	−94999.978	−93038.260

				High speededness level (HSL)	
	Fitted model			**Mixture model with SC**	**Non-mixture model with SC**
	Non-mixture model with SC ($\rho_{\theta\tau} = 0.8$)	DIC	Q_1	191396.999	187321.113
			Median	191702.770	187686.570
True Model			Q_3	192171.363	187941.382
		LPML	Q_1	−95202.124	−93221.728
			Median	−95101.015	−93157.626
			Q_3	−95012.373	−93028.099

Note that the mixture model is the model in Section 2. The non-mixture model is the hierarchical structure model in van der Linden (2007).

the non-mixture model (−103871.796). Similarly, under the HSL and SC conditions, when the mixture model was the true model, the mixture model also fitted the data best. The differences in the medians of DIC and LPML between the mixture model and non-mixture model were −16705.267 and 6977.566, respectively. In addition, under the LSL and WC conditions,

when the non-mixture model was the true model, the non-mixture model fitted the data better. The median DIC of the non-mixture model (188057.725) was smaller than that of the mixture model (192051.824), and the median LPML of the non-mixture model (−93222.498) was larger than that of the mixture model (−95204.235). Similarly, under the HSL and SC conditions, when
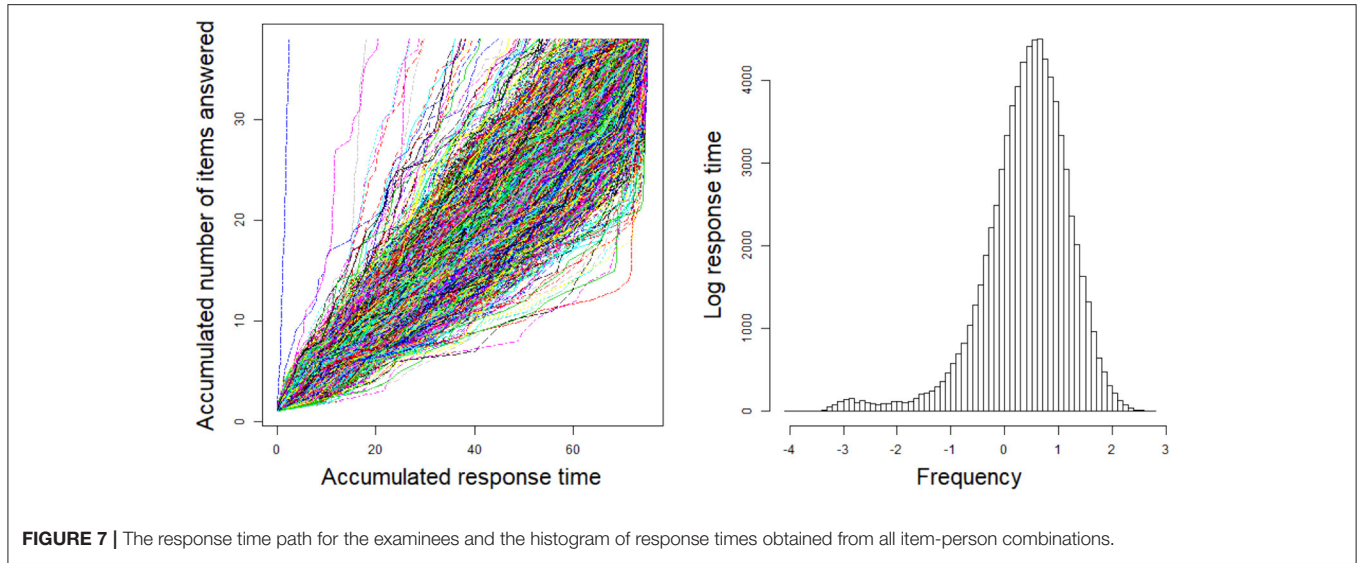
FIGURE 7 | The response time path for the examinees and the histogram of response times obtained from all item-person combinations.

TABLE 5 | The estimation results of discrimination and difficulty parameter for the real data.

Para.		EAP		SD		HPDI	
a	b	$\hat{a}$	$\hat{b}$	SD_a	SD_b	$HPDI_a$	$HPDI_b$
a_1	b_1	0.8182	−0.8649	0.0009	0.0008	[0.7591, 0.8832]	[−0.9234, −0.8073]
a_2	b_2	0.7302	−1.1924	0.0007	0.0015	[0.6809, 0.7862]	[−1.2680, −1.1134]
a_3	b_3	0.4409	−1.2129	0.0003	0.0028	[0.4034, 0.4786]	[−1.3152, −1.1096]
a_4	b_4	0.2000	−1.0279	0.0000	0.0030	[0.1863, 0.2036]	[−1.1353, −0.9183]
a_5	b_5	0.6192	−0.7536	0.0007	0.0010	[0.5652, 0.6715]	[−0.8159, −0.6888]
a_6	b_6	0.5618	−1.0134	0.0005	0.0016	[0.5150, 0.6075]	[−1.0982, −0.9389]
a_7	b_7	0.6946	−1.8531	0.0005	0.0027	[0.6518, 0.7405]	[−1.9656, −1.7591]
a_8	b_8	0.3710	−1.3215	0.0003	0.0042	[0.3350, 0.4046]	[−1.4438, −1.1925]
a_9	b_9	0.5969	−0.6650	0.0008	0.0010	[0.5441, 0.6552]	[−0.7280, −0.6072]
a_{10}	b_{10}	0.6228	−0.9849	0.0007	0.0015	[0.5738, 0.6769]	[−1.0609, −0.9129]
a_{11}	b_{11}	0.5124	−0.1673	0.0008	0.0004	[0.4601, 0.5719]	[−0.2073, −0.1293]
a_{12}	b_{12}	0.7251	−0.8260	0.0008	0.0009	[0.6674, 0.7812]	[−0.8851, −0.7662]
a_{13}	b_{13}	0.3342	−1.5034	0.0002	0.0058	[0.3011, 0.3663]	[−1.6613, −1.3594]
a_{14}	b_{14}	0.5786	−0.0406	0.0008	0.0003	[0.5179, 0.6319]	[−0.0784, −0.0093]
a_{15}	b_{15}	0.3464	−1.2434	0.0003	0.0045	[0.3140, 0.3846]	[−1.3769, −1.1141]
a_{16}	b_{16}	1.0816	−0.8625	0.0006	0.0006	[1.0050, 1.1628]	[−0.9109, −0.8092]
a_{17}	b_{17}	0.4434	−1.3966	0.0003	0.0035	[0.4070, 0.4823]	[−1.5151, −1.2828]
a_{18}	b_{18}	0.6631	−0.2462	0.0010	0.0003	[0.6023, 0.7263]	[−0.2826, −0.2071]
a_{19}	b_{19}	0.5072	−0.8406	0.0005	0.0015	[0.4600, 0.5525]	[−0.9186, −0.7620]
a_{20}	b_{20}	0.2638	−0.7837	0.0003	0.0042	[0.2251, 0.2972]	[−0.9173, −0.6637]
a_{21}	b_{21}	0.5548	−0.7497	0.0006	0.0012	[0.5030, 0.6056]	[−0.8212, −0.6832]
a_{22}	b_{22}	0.6791	−0.4723	0.0010	0.0006	[0.6150, 0.7403]	[−0.5235, −0.4273]
a_{23}	b_{23}	0.4225	−0.7727	0.0005	0.0019	[0.3803, 0.4670]	[−0.8579, −0.6881]
a_{24}	b_{24}	0.7590	−0.5959	0.0011	0.0006	[0.6925, 0.8225]	[−0.6477, −0.5447]
a_{25}	b_{25}	0.8798	−0.6894	0.0012	0.0006	[0.8136, 0.9525]	[−0.7414, −0.6393]
a_{26}	b_{26}	0.7344	−0.4227	0.0011	0.0005	[0.6680, 0.7990]	[−0.4683, −0.3774]
a_{27}	b_{27}	0.5176	−0.6252	0.0007	0.0013	[0.4685, 0.5720]	[−0.6943, −0.5492]
a_{28}	b_{28}	0.7185	−0.7225	0.0009	0.0009	[0.6601, 0.7822]	[−0.7846, −0.6619]

(Continued)

TABLE 5 | Continued

Para.		EAP		SD		HPDI	
a	b	$\hat{a}$	$\hat{b}$	SD_a	SD_b	$HPDI_a$	$HPDI_b$
a_{29}	b_{29}	0.7444	−0.7613	0.0009	0.0009	[0.6797, 0.8024]	[−0.8245, −0.7029]
a_{30}	b_{30}	0.5110	−0.4083	0.0007	0.0008	[0.4550, 0.5658]	[−0.4709, −0.3542]
a_{31}	b_{31}	0.4307	−0.0292	0.0007	0.0009	[0.3775, 0.4843]	[−0.0911, 0.0303]
a_{32}	b_{32}	0.7277	−0.4895	0.0011	0.0008	[0.6624, 0.7954]	[−0.5451, −0.4327]
a_{33}	b_{33}	0.5667	0.0485	0.0009	0.0004	[0.5097, 0.6253]	[0.0035, 0.0905]
a_{34}	b_{34}	0.2024	−0.7727	0.0000	0.0067	[0.2000, 0.2152]	[−0.9325, −0.6074]
a_{35}	b_{35}	0.6925	−0.6144	0.0012	0.0029	[0.6239, 0.7624]	[−0.7182, −0.5086]
a_{36}	b_{36}	0.6983	−0.2498	0.0014	0.0064	[0.6228, 0.7744]	[−0.3874, −0.0890]
a_{37}	b_{37}	0.4374	0.4227	0.0017	0.0097	[0.3525, 0.5189]	[0.1555, 0.6958]

Para. denotes the interest parameters. EAP denotes the expected a priori estimation. SD denotes the standard deviation. HPDI denotes the 95% highest posterior density intervals.

the non-mixture model was the true model, the mixture model also fitted the data better. The differences in the medians of DIC and LPML between the non-mixture model and mixture model were −4016.200 and 1943.389, respectively. Refer to **Tables 3**, **4** for more detailed results of the model assessment. In summary, the Bayesian assessment criteria were effective for identifying the true models and could, thus, be used in the subsequent real data study.

6. EMPIRICAL EXAMPLE

This section presents an application of the mixture model with an empirical example. The data set was from a high-state, large-scale, standardized computerized adaptive test that was previously analyzed by Wang and Xu (2015). The data set included 37 dichotomous items, and the test time was 75 min. The sample size was 2,106. The mixture model and non-mixture model were used to fit the item response and response time data of the 37 dichotomous items. The response time path for the examinees is shown in **Figure 7**. In addition, **Figure 7** shows a histogram of response times obtained from all item–person combinations.

In the Bayesian computation, we used 20,000 MCMC samples after a burn-in of 10,000 iterations to compute all posterior estimates. The convergence of the chains was checked using the PSRF. The PSRF values of all item parameters were less than 1.2. We used the DIC and LPML to fit the mixture model and non-mixture model. The mixture model resulted in a smaller DIC value (350696.11) than the non-mixture model (365690.66), and the LPML of the mixture model (−175027.99) was larger than that of the non-mixture model (−181062.48). This indicates that the mixture model better fitted the data. Based on the results of the model assessment, we used the mixture model to analyze real data in detail.

Analysis of item parameters

The estimated results for the discrimination and difficulty parameters are shown in **Table 5**. As shown in the table, the expected a posteriori (EAP) estimates of the one-item

discrimination parameters were greater than 1. This indicated that the items could well distinguish the differences between abilities. The three items with the lowest discrimination were items 4, 34, and 20. The EAP estimates of discrimination parameters for these three items were 0.2000, 0.2024, and 0.2638. In addition, another three items had the lowest EAP estimates of the difficulty parameters, indicating that these items were easier than the other items. These were items 7, 13, and 8. The EAP estimates of g_j had a range of 0.1334 to 0.2945. The EAP estimates of λ_j had a range of −0.3322 to 0.7634.

7. CONCLUSION

In this article, we propose a novel and efficient Bayesian algorithm (Pólya–gamma Gibbs sampling algorithm) based on the auxiliary variables for estimating the mixture hierarchical model. The new algorithm avoids the tedious multidimensional integral operation of the MMLE. Within a fully Bayesian framework, the Pólya–gamma Gibbs sampling algorithm not only avoids the heavy reliance of the traditional Metropolis–Hastings algorithm on the tuning parameters of the proposed distributions for different data sets but also overcomes the disadvantage of the Metropolis–Hastings algorithm being sensitive to step size. However, the computational burden of the Pólya–gamma Gibbs sampling algorithm becomes excessive especially when there are a large number of examinees, the items or the abnormal response and response time data are considered, or a large number MCMC sample size is used. Therefore, it would be desirable to develop a stand-alone R package associated with Fortran software for a more extensive large-scale assessment program.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.841372/full#supplementary-material

REFERENCES

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibb sampling. *J. Educ. Stat.* 17, 251–269.

Albert, J. H., and S. Chib. (1993). Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* 88, 669–679.

Asparouhov, T., and Muthén, B. (2010b). *Bayesian analysis of latent variable models using Mplus* (Technical report, Version 4). Available online at: http://www.statmodel.com.

Baker, F. B., and Kim, S. H. (2004). *Item Response Theory: Parameter Estimation Techniques.* 2nd Edition, Boca Raton: CRC Press. doi: 10.1201/9781482276725

Béguin, A. A., and Glas, C. A. W. (2001). MCMC estimation of multidimensional IRT models. *Psychometrika* 66, 541–561. doi: 10.1007/BF02296195

Biane, P., Pitman, J., and Yor, M. (2001). Probability laws related to the Jacobi theta and Riemann zeta functions, and brownian excursions. *Bull. Am. Math. Soc.* 38, 435–465. doi: 10.48550/arXiv.math/9912170

Birnbaum, A. (1968). "Some latent trait models and their use in inferring an examinee's ability," in *Statistical Theories of Mental Test Scores*, eds F. M. Lord and M. R. Novick (Reading, MA: Addison-Wesley), 397–472.

Bock, R. D., and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika* 46, 443–459.

Bock, R. D., and Schilling, S. G. (1997). "High-dimensional full-information item factor analysis," in *Latent Variable Modelling and Applications to Causality*, ed M. Berkane (New York, NY: Springer), 164–176.

Bolt, D. M., Cohen, A. S., and Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: application of a mixture Rasch model with ordinal constraints. *J. Educ. Meas.* 39, 331–348. doi: 10.1111/j.1745-3984.2002.tb01146.x

Boughton, K. A., and Yamamoto, K. (2007). "A HYBRID model for test speededness," in *Multivariate and Mixture Distribution Rasch Models*, eds M. von Davier and C. H. Carstensen (New York, NY: Springer), 147–156.

Brooks, S. P., and Gelman, A. (1998). Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.

Chang, Y. W., Tsai, R. C., and Hsu, N. J. (2014). A speeded item response model: leave the harder till later. *Psychometrika* 79, 255–274. doi: 10.1007/s11336-013-9336-2

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York, NY: Springer.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *Am. Stat.* 49, 327–335.

Converse, G., Curi, M., Oliveira, S., and Templin, J. (2021). Estimation of multidimensional item response theory models with correlated latent variables using variational autoencoders. *Mach. Learn.* 110, 1463–1480. doi: 10.1007/s10994-021-06005-7

Fox, J.-P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. New York, NY: Springer.

Fox, J.-P., and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika* 66, 269–286. doi: 10.1007/BF02294839

Fox, J. P. (2005). Multilevel IRT using dichotomous and polytomous items. *Br. J. Math. Stat. Psychol.* 58, 145–172. doi: 10.1348/000711005X38951

Geisser, S., and Eddy, W. F. (1979). A predictive approach to model selection. *J. Am. Stat. Assoc.* 74, 153–160.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). "Model determination using predictive distributions with implementation *via* sampling-based methods (with discussion)," in *Bayesian statistics 4*, eds J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Oxford, UK: Oxford University Press), 147–167.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85, 398–409.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statisti Sci*, 7, 457–472. doi: 10.1214/ss/1177011136

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 721–741.

Ghosh, M., A. Ghosh, M. Chen, and A. Agresti. (2000). Noninformative priors for one parameter item response models. *Journal of Statistical Planning and Inference*, 88, 99-115.

Goegebeur, Y., De Boeck, P., Wollack, J. A., and Cohen, A. S. (2008). A speeded item response model with gradual process change. *Psychometrika* 73, 65–87. doi: 10.1007/s11336-007-9031-2

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York, NY: Springer.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Chichester, UK: John Wiley & Sons.

Jiang, Z., and Templin, J. (2019). Gibbs samplers for logistic item response models *via* the pólya–gamma distribution: a computationally efficient data–augmentation strategy. *Psychometrika* 84, 358?374. doi: 10.1007/s11336-018-9641-x

Kuk, A. Y. C. (1999). Laplace importance sampling for generalized linear mixed models. *J. Stat. Comput. Simulat.* 63, 143–158.

Lee, S.-Y., and Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behav. Res.* 39, 653–686. doi: 10.1207/s15327906mbr3904_4

Lord, F. M., and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley.

Lord, F. M., and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Man, K., and Harring, J. R. (2021). Assessing preknowledge cheating *via* innovative measures: A multiple-group analysis of jointly modeling item responses, response times, and visual fixation counts. *Educ. Psychol. Meas.* 81, 441–465. doi: 10.1177/0013164420968630

Man, K., Harring, J. R., Ouyang, Y., and Thomas, S. L. (2018). Response time based nonparametric Kullback-Leibler divergence measure for detecting aberrant test-taking behavior. *Int. J. Testing* 18, 155–177. doi: 10.1080/15305058.2018.1429446

Maris, E. (1993). Additive and multiplicative models for gamma distributed variables, and their application as psychometric models for response times. *Psychometrika* 58, 445–469.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chemi. Phys.* 21, 1087–1092.

Muthén, B. O. (2010). *Bayesian Analysis in Mplus: A Brief Introduction (Incomplete Draft,Version 3).* Available online at: http://www.statmodel.com/download/IntroBayesVersion%203.pdf.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Am. Stat. Assoc.* 108, 1339–1349. doi: 10.1080/01621459.2013.829001

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2002). Reliable estimation of general ized linear mixed models using adaptive quadrature. *Stata J.* 2, 1–21. doi: 10.1177/1536867X0200200101

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *J. Econometr.* 128, 301–323. doi: 10.1016/j.jeconom.2004.08.017

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., and Heathcote, A. (2015). The lognormal race: a cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika* 80, 491–513. doi: 10.1007/s11336-013-9396-3

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., and Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika* 68, 589–606. doi: 10.1007/BF02295614

Scheiblechner, H. (1979). Specific objective stochastic latency mechanisms. *J. Math. Psychol.* 19, 18–38.

Schnipke, D. L., and Scrams, D. J. (1997). Modeling response times with a two-state mixture model: a new method of measuring speededness. *J. Educ. Meas.* 34, 213–232.

Skaug, H. J. (2002). Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models. *J. Comput. Graphical Stat.* 11, 458–470. doi: 10.1198/106186002760180617

Song, X.-Y., and Lee, S.-Y. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *J. Math. Psychol.* 56, 135–148. doi: 10.1016/j.jmp.2012.02.001

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B* 64, 583–639. doi: 10.1111/1467-9868.00353

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* 82, 528–550.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussions). *Ann. Stat.* 22, 1701–1762.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *J. Educ. Behav. Stat.* 31, 181–204. doi: 10.3102/10769986031002181

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika* 72, 287–308. doi: 10.1007/s11336-006-1478-z

van der Linden, W. J., and Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika* 73, 365–384. doi: 10.1007/s11336-007-9046-8

Wang, C., Fan, Z., Chang, H.-H., and Douglas, J. (2013). A semiparametric model for jointly analyzing response times and accuracy in computerized testing. *J. Educ. Behav. Stat.* 38, 381–417. doi: 10.3102/1076998612461831

Wang, C., and Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *Br. J. Math. Stat. Psychol.* 68, 456–477. doi: 10.1111/bmsp.12054

Wang, C., Xu, G., and Shang, Z. (2018). A two-stage approach to differentiating normal and aberrant behavior in computer based testing. *Psychometrika* 83, 223–254. doi: 10.1007/s11336-016-9525-x

Wise, S. L., and DeMars, C. E. (2006). An application of item response time: the effort-moderated IRT model. *J. Educ. Meas.* 43, 19–38. doi: 10.1111/j.1745-3984.2006.00002.x

Wise, S. L., and Kong, X. (2005). Response time effort: a new measure of examinee motivation in computer-based tests. *Appl. Meas. Educ.* 18, 163–183. doi: 10.1207/s15324818ame1802_2

Zhang, J. W., Lu, J., Chen, F., and Tao, J. (2019). Exploring the correlation between multiple latent variable and covariates in hierarchical data based on the multilevel multidimensional IRT model. *Front Psychol*, 10:2387. doi: 10.3389/fpsyg.2019.02387

Zhang, Z., Y., Zhang, J. W., Lu, J., and Tao, J. (2020). Bayesian estimation of the DINA model with Pólya-Gamma Gibbs algorithm. *Front. Psychol.* 11, 384. doi: 10.3389/fpsyg.2020.00384

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.