Psychometrika (2026), **91**, 1280–1321
doi:10.1017/psy.2026.10121

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Bayesian Modeling and Inference for Item Response Model with Nonignorable Missing Data

Jing Wu[1], Zhihua Ma[2] and Ming-Hui Chen[3]

[1]Computer Science and Statistics, University of Rhode Island, USA; [2]Statistics, Shenzhen University, China; [3]Statistics, University of Connecticut, USA

Corresponding author: Ming-Hui Chen; E-mail: ming-hui.chen@uconn.edu

(Received 7 May 2025; revised 20 April 2026; accepted 27 April 2026)

Abstract

Not-reached (dropout) and omitted (intermittent missingness) behaviors are often inevitable in timed computerized tests. These missingness behaviors may be related to the subject's latent traits, the difficulty of the item, or even the unobserved item response itself. In order to better understand the underlying test-taking behaviors, a Bayesian hierarchical framework is adopted to jointly model the item response, the item response time, the not-reached, and the omitted behaviors. For missing data, a sequential multinomial model is developed for the not-reached behavior, and a conditional model is proposed for the omitted behavior conditioning on the not-reached behavior. A decomposed logarithm of the pseudo marginal likelihood (LPML) is then developed to assess the fit of the missing data models. It can be further used to quantify the importance of modeling item response and response time jointly versus individually in identifying the missing data mechanism. The empirical performance of the proposed models and the model assessment criterion are examined through extensive simulations. The proposed methodology is further applied to analyze the Program for International Student Assessment (PISA) 2018 Test.

Keywords: dropout; intermittent missingness; IRT; nonignorable; omitted item

1. Introduction

Missing data are frequently encountered in the fields of education, psychology, and social science due to the nature of data collection through tests or questionnaires, where responses are often not mandatory (Rose et al., 2017; Yuan et al., 2018). In timed tests, two types of missing behaviors can occur. One is called not-reached or dropout (Wu et al., 2019), where test takers fail to reach the end of the test due to time limits, early quitting, strategic planning, or other reasons. The other type is called omitted or intermittent missingness, where test takers skip questions intentionally or accidentally, or run out of time due to item-level time limits. For example, in the 2018 PISA math exam data in Brazil, 18.7% of test takers did not complete the test and 18.6% of answers was missing. Ignoring or handling missing data inappropriately can result in loss of information or even bias when analyzing the testing results (Glas & Pimentel, 2008; Rose et al., 2015).

To appropriately address the missing data problem, one needs to understand different types of missing data mechanisms and their underlying assumptions. The term "ignorable" missingness was first introduced by Rubin (1976), meaning that the inference for the response models is still valid if

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution-ShareAlike licence (https://creativecommons.org/licenses/by-sa/4.0), which permits re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

the missing data process is ignored. In contrast, missing data are considered nonignorable if Rubin's assumption of ignorable missingness is violated (Little & Rubin, 2019; Rabe-Hesketh & Skrondal, 2023). According to Glas and Pimentel (2008), Guo et al. (2022), and Rose et al. (2017), the not-reached and omitted behaviors may depend on the latent characteristics and the difficulty of the items and thus cannot be ignored for valid inference. The missingness is thus more likely to be nonignorable. Furthermore, these two behaviors are different processes that result in item nonresponses with different characteristics, and therefore should be handled differently in the item response theory (IRT) models (Rose, 2013; Rose et al., 2017).

Approaches for dealing with the aforementioned nonignorable missingness for item responses have been developed. O'Muircheartaigh and Moustaki (1999) proposed a shared parameter joint model for item responses and missing data indicators, where the tendency to omit items is included as a latent variable shared by both models. Following this framework, Holman and Glas (2005) considered a general IRT model framework via a combination of a generalized partial credit model (GPCM) for the observed dichotomous and polytomous responses and a multidimensional IRT (MIRT) model for the missing data process. Glas et al. (2015) generalized the approach to incorporate covariates into the joint model of latent variables. For all the aforementioned papers, they focused only on missingness due to omission. Following the modeling framework of Holman and Glas (2005), Glas and Pimentel (2008) assumed a sequential IRT model with linear restrictions for the missing data model of the not-reached behavior. List et al. (2019) extended the model by applying a semiparametric approach to capture nonlinear relationships between the onset of the not-reached items and the characteristics of test takers. Both papers focused on the unobserved responses at the end of the response sequence (the not-reached item) and did not consider missingness due to omission. To account for both omitted and not-reached missing behaviors, Rose et al. (2017) proposed a joint model that combines an MIRT model for response and omitted items, and a latent regression model (LRM) for latent variables with a function of the number of not-reached items as a predictor. Consequently, the not-reached behavior of test takers was not directly modeled. Guo (2019) and Zhang et al. (2025) proposed a selection model to jointly study the item response and the missing data, where the cumulative number of missing indicators and cumulative responses were used as covariates in the missing data model. Our model adopts the joint modeling philosophy, but extends it by differentiating between intermittent and not-reached missingness to account for their distinct underlying causes, and incorporating response time data to enhance the modeling of item response and missing data process.

The response time for each item can sometimes be easily collected because of the prevalence of computerized testing. Many recent developments incorporate this additional source of information in the analysis of missing item response data, with the hope of improving the estimation and inference of item parameters and subjects' latent characteristics (Guo et al., 2022; Lu & Wang, 2020; Lu et al., 2023; Pohl et al., 2019; Ulitzsch et al., 2020c; Van der Linden, 2007). Pohl et al. (2019) adopted a hierarchical framework to jointly model response accuracy and response time, assuming that the missingness was caused solely by time limits, an assumption not applicable to our real data, where the median total time spent on the test for not-reached subjects (23.26 minutes) is far less than the test time limit (60 minutes). Ulitzsch et al. (2020c) adopted the speed and accuracy model proposed in Van der Linden (2007) and the IRT model for the omitted behavior proposed in Holman and Glas (2005), and assumed different response time models for the observed and omitted items. Ulitzsch et al. (2020a) further included the modeling of engagement (guessing) behavior in the hierarchical model framework, and assumed fully engaged students did not skip questions. Both papers focused exclusively on omission behavior. Lu and Wang (2020) proposed a response time process (RTprocess) model where they modeled item omission as a type of censoring. They assumed that the omitted items were due to item-level time limits, and that the not-reached items were not the result of early quitting. Guo et al. (2022) proposed a joint model for responses and response times that accounts for missingness due to the not-reached behavior, where a censoring mechanism was assumed for the response time. Similarly, the not-reached and omitted behaviors were not directly modeled. Lu et al. (2023) proposed a two-stage mixture RTprocess model to detect aberrant behaviors (i.e., cheating and guessing behaviors) and

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

to account for item nonresponses (not-reached and omitted items). Their model did not differentiate between the not-reached and omitted behaviors and assumed that the missing data mechanism was ignorable if the response time information was incorporated. Ulitzsch et al. (2020b, 2025) proposed a model that accounts for different types of not-reached behaviors: a response time model to address not-reached behavior due to time limits, and a Poisson model to address not-reached behavior due to early quitting. Their paper centered solely on not-reached behavior. Ulitzsch et al. (2024) distinguished rapid omissions from omissions due to solution behavior and considered omission classification uncertainty. Their paper focused only on missingness due to omission. Guo et al. (2025) proposed a joint model of item response, response time, and omitted time through the competing risk formulation. The three outcomes were assumed to be conditionally independent given the latent traits, where a multivariate normal distribution was adopted. Their paper focused on missingness due to omission only.

Not-reached and omitted behaviors often stem from different behavioral processes, such as time limits, early quitting, or strategic planning, which may follow distinct missing data mechanisms. These two types of behaviors should be modeled separately to capture their unique characteristics, but also jointly, as they may both be related to the same underlying latent ability and item difficulty parameters (Debeer et al., 2017; Mislevy & Wu, 1988). Joint modeling of both behaviors allows for a more accurate understanding of test-taking behavior and results in improved parameter estimation by borrowing information between both types of missing behaviors. Guo et al. (2025) proposed a joint model framework for both omitted items and not-reached items by combining the competing risk models from Guo et al. (2025) and the censoring approach introduced in Guo et al. (2022). Their focus was on missingness resulting from time limits. The existing literature that simultaneously modeling both not-reached and omitted behaviors is sparse, and our study aims to fill this gap. In this article, we propose a joint model framework to simultaneously study the item response, response time, the not-reached missingness, and the omitted missingness, without restricting missingness to time constraints. Specifically, a summary of the contributions of this work is listed below.

(i) A data-driven missing data model is developed to allow for joint modeling of the not-reached and omitted behaviors of test takers. Specifically, a sequential multinomial model is developed for the not-reached behavior, and a conditional model is proposed for the omitted behavior given the not-reached behavior. This formulation enables the modeling of not-reached and omission behaviors within a unified IRT modeling framework. It also provides a clear and natural distinction between the two types of missing behaviors: the omitted behavior, which may be due to strategic planning, and the not-reached behavior, which may be due to early quitting or time limits. Our proposed missing data model assumes nonignorable because the probability of missingness does not directly depend on any unobserved data but only depends on latent variables, and the missing data process cannot be ignored for valid inference of the response models (Pohl & Becker, 2020; Robitzsch, 2021, 2023). The model also encompasses an ignorable missing data mechanism as a special case.

(ii) We propose a hierarchical model framework and a new formulation for the joint distributions of all latent propensity parameters. This allows us to quantify the extent to which the missing data mechanism is nonignorable and to conveniently incorporate the conditional independence properties of latent propensity parameters, facilitating the specification of prior distributions more conveniently and resulting in a more efficient implementation and better mixing of Markov chain Monte Carlo (MCMC) sampling. Our simulation studies demonstrate good empirical performances even when the missing data percentages are as small as 5%, either due to omission or due to not-reached behaviors.

(iii) The decomposition techniques of various model selection criteria have been widely applied to the IRT to enhance the understanding of model fit within the joint model framework (Liu et al., 2025; Sun et al., 2025). In this article, we also develop a model selection criterion, built on the decomposition of the logarithm of the pseudo marginal likelihood (LPML), a

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

computationally efficient posterior predictive criterion, using the idea of leave-one-subject-out. The decomposition of LPML gatekeeps the selection of the "best" missing data mechanism. More importantly, it can also be used to quantify the importance of modeling the item response and response time jointly versus individually in identifying the missing data mechanism.

The remainder of this article is organized as follows. We present models for the item response, response time, and missing data mechanism in Section 2. The hierarchical model framework and Bayesian inference are given in Sections 3 and 4, including the decomposition of LPML for model fit assessment. Extensive simulation studies and a detailed analysis of the PISA 2018 data are presented in Sections 5 and 6, respectively. Finally, we conclude this article with a brief discussion in Section 7.

2. Models

2.1. Item response model

Suppose that there are a total of n subjects and J items. Let y_{ij} denote the binary score for the i^{th} subject at item j, where $y_{ij} = 1$ indicates that the answer is correct and $y_{ij} = 0$ otherwise. Denoting $P(y_{ij} = 1) = p_{y_{ij}}$, a two-parameter logistic model (2PL) (Bock & Aitkin, 1981) is assumed for $p_{y_{ij}}$ as

$$p_{y_{ij}} = \frac{\exp(a_{Yj}(\theta_{Yi} - b_{Yj}))}{1 + \exp(a_{Yj}(\theta_{Yi} - b_{Yj}))}, \tag{1}$$

for $i = 1, \ldots, n, j = 1, \ldots, J$, where $\mathbf{a}_Y = (a_{Y1}, \ldots, a_{YJ})'$ is the vector of discrimination parameters, $\boldsymbol{\theta}_Y = (\theta_{Y1}, \ldots, \theta_{Yn})'$ is the vector of latent response propensity parameters (O'Muircheartaigh & Moustaki, 1999) measuring the tendency to give correct answers, and $\mathbf{b}_Y = (b_{Y1}, \ldots, b_{YJ})'$ is the vector of difficulty parameters. A higher value of the product of a_{Yj} and b_{Yj} indicates that the corresponding item j is more difficult and, consequently, the answer is more likely to be wrong.

2.2. Response time model

Let t_{ij} denote the response time for the i^{th} subject at item j. A log-normal model (Van der Linden, 2006, 2007) is built as

$$\log(t_{ij}) = b_{Tj} - \sigma_T \theta_{Ti} + \varepsilon_{ij}, \tag{2}$$

where $\mathbf{b}_T = (b_{T1}, \ldots, b_{TJ})'$ is the vector of time intensity parameters, $\theta_T = (\theta_{T1}, \ldots, \theta_{Tn})'$ is the vector of latent response time propensity parameters, $\sigma_T > 0$ captures the subject-specific variation, $\varepsilon_{ij} \sim N(0, \frac{1}{a_{Tj}})$, and $\mathbf{a}_T = (a_{T1}, \ldots, a_{TJ})'$ is the vector of precision parameters. A higher value of b_{Tj} indicates that subjects tend to take more time on item j.

2.3. Missing data models

We adopt a sequential multinomial model to study the not-reached behavior of each subject. A joint model of all the previous missing statuses, that is, omitted behavior, is then developed given that the subject fails to reach a certain item.

Let W_i denote the item number at which the i^{th} subject fails to reach, for $i = 1, \ldots, n$. The possible values of W_i are $1, 2, \ldots, J, J+1$, where $W_i = J+1$ indicates that the i^{th} subject reaches the last item.

Let

$$p_{ij} = P(W_i = j | j \le W_i \le J+1) \text{ if } 1 \le j \le J. \tag{3}$$

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

We then have

$$P(W_i = j) = \begin{cases} p_{i1} & \text{if } j = 1, \\ \prod_{\ell=1}^{j-1}(1-p_{i\ell})p_{ij} & \text{if } 1 < j \le J, \\ \prod_{\ell=1}^{J}(1-p_{i\ell}) & \text{if } j = J+1, \end{cases} \tag{4}$$

A Rasch model (Rasch, 1993) is assumed for p_{ij} in (3):

$$p_{ij} = \frac{\exp(\theta_{Wi} - b_{Wj})}{1+\exp(\theta_{Wi} - b_{Wj})}, \tag{5}$$

for $i = 1,\ldots,n, j = 1,\ldots,J$. Let $\theta_W = (\theta_{W1},\ldots,\theta_{Wn})'$ denote the vector of latent not-reached propensity parameters (O'Muircheartaigh & Moustaki, 1999; Ulitzsch et al., 2020a) that measure the tendency to withdraw from the test and $\mathbf{b}_W = (b_{W1},\ldots,b_{WJ})'$ denote the vector of difficulty parameters for the not-reached behavior. This allows us to model the not-reached behavior within a unified IRT modeling framework.

Let $R_{i\ell}$ be the binary variable that indicates whether item ℓ is omitted or not for the i^{th} subject:

$$R_{i\ell} = \begin{cases} 0 & y_{i\ell} \text{ is observed}, \\ 1 & y_{i\ell} \text{ is missing}. \end{cases}$$

Assuming that the i^{th} subject fails to reach item j ($W_i = j$), we know for sure that the missing indicator $R_{i\ell} = 1$ with probability one for $\ell \ge j$. Moreover, $y_{i,j-1}$ must be observed ($R_{i,j-1} = 0$) with probability one. Otherwise, $W_i < j$, which contradicts the condition that $W_i = j$. Thus, given $W_i = j$, we only need to model the conditional joint distribution of $R_{i\ell}$ up to $\ell = j-2$.

Following Ibrahim, Chen, and Lipsitz (2001) and Ibrahim et al. (2005), we write the joint distribution of $(R_{i1},\ldots,R_{ij-2})$ through a sequence of one-dimensional conditional distributions. Specifically, for $j = 2$, $P(R_{i1} = r_{i1}|W_i = 2) = 1 - r_{i1}$; for $j = 3$, $P(R_{i1} = r_{i1}, R_{i2} = r_{i2}|W_i = 3) = f(r_{i1}|W_i = 3)(1 - r_{i2})$; and for $j > 3$, $P(R_{i1} = r_{i1},\ldots,R_{i,j-2} = r_{i,j-2}|W_i = j) = f(r_{i1}|W_i = j)\prod_{\ell=2}^{j-2} f(r_{i\ell}|W_i = j, r_{i1},\ldots,r_{i,\ell-1})$. We assume the following formulation for the conditional distribution of $R_{i\ell}$ given $W_i = j$ and the previous omitted indicators:

$$\begin{aligned} &P(R_{i\ell} = 1|W_i = j, R_{i1} = r_{i1},\ldots,R_{i,\ell-1} = r_{i,\ell-1}) \\ &= P(R_{i\ell} = 1|R_{i1} = r_{i1},\ldots,R_{i,\ell-1} = r_{i,\ell-1}) \\ &= \begin{cases} q_{i\ell} & \text{if } 1 \le \ell \le j-2, \\ 0 & \text{if } \ell = j-1, \end{cases} \end{aligned} \tag{6}$$

where $q_{i\ell}$ is modeled by another Rasch model:

$$q_{i\ell} = \frac{\exp(\theta_{Ri} - b_{R\ell})}{1+\exp(\theta_{Ri} - b_{R\ell})}, \tag{7}$$

for $i = 1,\ldots,n, \ell = 1,\ldots,j-2$. Let $\theta_R = (\theta_{R1},\ldots,\theta_{Rn})'$ denote the vector of latent omission propensity parameters that measures the tendency to omit the item, and $\mathbf{b}_R = (b_{R1},\ldots,b_{R,J-1})'$ denote the vector of difficulty parameters for the omitted behavior model. A higher value of $b_{R\ell}$ indicates that the corresponding item ℓ is less difficult and tends to be answered.

The joint distribution of W_i and $\mathbf{R}_i = (R_{i1},\ldots,R_{iJ})'$ can be written as

$$\begin{aligned} &P(W_i = w_i, R_{i1} = r_{i1},\ldots,R_{iw_i} = r_{iw_i}) \\ &= \left\{\prod_{\ell=1}^{w_i-1}(1-p_{i\ell})\right\}^{\mathbf{1}(1<w_i\le J+1)} p_{iw_i}{}^{\mathbf{1}(w_i<J+1)} \left\{\prod_{j=1}^{w_i-2} q_{ij}^{r_{ij}}(1-q_{ij})^{1-r_{ij}}\right\}^{\mathbf{1}(w_i>1)}. \end{aligned}$$

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Remark 2.1. The sequential model framework in (3) and (4) allows us to directly interpret $p_{ij} = P(W_i = j | j \leq W_i < J+1)$, that is, the chance of failing to reach the current item given that the subject has not withdrawn from the test. This is inspired by the survival function in survival analysis, where p_{ij} is of great practical interest and should not depend on later items. The sequential model framework implicitly takes into account the order of items, where each p_{ij} depends on the item values only up to j, rather than taking them as a whole in a conventional multinomial logistic regression model (Wu et al., 2019). This aligns well with the design of certain computer-based tests, such as PISA 2018, which do not allow students to return to a question once it has been answered or skipped. The proposed sequential logistic regression model is therefore more appealing. The sequential multinomial model for the not-reached behavior automatically sets the upper limit of the total number of not-reached items and allows for incorporating both subject-specific and item-specific parameters, which is expected to provide a better fit for the data than a more parsimonious Poisson model in Ulitzsch et al. (2020b, 2025).

Remark 2.2. The minimum requirements for the missing data models in (5) and (7) to be estimated are that i) the number of subjects failing to reach item j should be greater than 0 for any j and ii) the number of omissions should be greater than 0 for any item level. These conditions are akin to the traditional requirement of "adequate item response variability" in response models (Hambleton et al., 1991). In Section 3, we introduce a hierarchical framework where these requirements can be relaxed by borrowing information from items with missing data. Apart from the traditional assumptions for IRT models (Fox, 2010), the only additional assumption introduced to simplify the proposed missing data model is given in Equation (6): $R_{i\ell}$ is independent of W_i given $(R_{i1}, \ldots, R_{i,\ell-1})$. This assumption can potentially be relaxed if there is evidence that $P(R_{i\ell} = 1 | W_i = j, R_{i1} = r_{i1}, \ldots, R_{i,\ell-1} = r_{i,\ell-1})$ varies between different item levels, with a sufficiently large percentage of missing data and a sufficiently large sample size n, and a sufficiently small number of items J. Additionally, we assume that $q_{i\ell}$ does not depend on any function of $R_{ij}, j = 1, \ldots, \ell-1$ for better identifiability. This assumption can also be relaxed with a large enough percentage of missing data and a large enough sample size (Köhler et al., 2015).

3. Hierarchical model framework

3.1. Latent propensity parameters

For the i^{th} subject, let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})'$ denote the vector of all response measures and $\mathbf{t}_i = (t_{i1}, \ldots, t_{iJ})'$ denote the vector of all response times. Let $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')'$, $\mathbf{t} = (\mathbf{t}_1', \ldots, \mathbf{t}_n')'$, $\mathbf{W} = (W_1, \ldots, W_n)'$, and $\mathbf{R} = (\mathbf{R}_1', \ldots, \mathbf{R}_n')'$. To capture the dependence between $\mathbf{y}$, $\mathbf{t}$, $\mathbf{W}$, and $\mathbf{R}$, we assume the following joint distributions of the latent propensity parameters:

$$(\theta_{Yi}, \theta_{Ti}, \theta_{Wi}, \theta_{Ri})' \sim N(\boldsymbol{\mu}, \Sigma), \tag{8}$$

where $\boldsymbol{\mu}$ is a four-dimensional column vector and Σ is the covariance matrix of $(\theta_{Yi}, \theta_{Ti}, \theta_{Wi}, \theta_{Ri})'$ given by

$$\Sigma = \begin{bmatrix} \sigma_\theta^2 & \sigma_{\theta\theta_T} & \sigma_{\theta\theta_W} & \sigma_{\theta\theta_R} \\ \sigma_{\theta\theta_T} & \sigma_{\theta_T}^2 & \sigma_{\theta_T\theta_W} & \sigma_{\theta_T\theta_R} \\ \sigma_{\theta\theta_W} & \sigma_{\theta_T\theta_W} & \sigma_{\theta_W}^2 & \sigma_{\theta_W\theta_R} \\ \sigma_{\theta\theta_R} & \sigma_{\theta_T\theta_R} & \sigma_{\theta_W\theta_R} & \sigma_{\theta_R}^2 \end{bmatrix}. \tag{9}$$

To ensure identifiability, let $\boldsymbol{\mu}$ be a four-dimensional column vector of zeros and the diagonal elements of Σ be 1.

Proposition 3.1. *For $i = 1, \ldots, n$, θ_{Wi} and θ_{Ri} are conditionally independent given θ_{Yi} and θ_{Ti} if and only if*

$$\sigma_{\theta_W\theta_R} - \frac{\sigma_{\theta\theta_W}\sigma_{\theta\theta_R}\sigma_{\theta_T}^2 - \sigma_{\theta\theta_W}\sigma_{\theta_T\theta_R}\sigma_{\theta\theta_T} - \sigma_{\theta\theta_R}\sigma_{\theta_T\theta_W}\sigma_{\theta\theta_T} + \sigma_{\theta_T\theta_R}\sigma_{\theta_T\theta_W}\sigma_\theta^2}{\sigma_\theta^2\sigma_{\theta_t}^2 - \sigma_{\theta\theta_t}^2} = 0. \tag{10}$$

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Proof of Proposition 3.1. For multivariate normal distributions, uncorrelated and independent properties are equivalent. Thus, we focus on the conditional covariance matrix of $(\theta_W, \theta_{Ri})'|(\theta_{Yi}, \theta_{Ti})'$, which is given by

$$\begin{aligned}\Sigma_{WR|YT} &= \begin{bmatrix} \sigma^2_{\theta_W} & \sigma_{\theta_W\theta_R} \\ \sigma_{\theta_W\theta_R} & \sigma^2_{\theta_R} \end{bmatrix} - \begin{bmatrix} \sigma_{\theta\theta_W} & \sigma_{\theta\theta_R} \\ \sigma_{\theta_T\theta_W} & \sigma_{\theta_T\theta_R} \end{bmatrix} \begin{bmatrix} \sigma^2_{\theta} & \sigma_{\theta\theta_T} \\ \sigma_{\theta\theta_T} & \sigma^2_{\theta_T} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{\theta\theta_W} & \sigma_{\theta_T\theta_W} \\ \sigma_{\theta\theta_R} & \sigma_{\theta_T\theta_R} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},\end{aligned}$$

where

$$\begin{aligned}\Sigma_{11} &= \sigma^2_{\theta_W} - \frac{\sigma^2_{\theta\theta_W}\sigma^2_{\theta_T} - 2\sigma_{\theta\theta_W}\sigma_{\theta_T\theta_W}\sigma_{\theta\theta_T} + \sigma^2_{\theta_T\theta_W}\sigma^2_{\theta}}{\sigma^2_{\theta}\sigma^2_{\theta_t} - \sigma^2_{\theta\theta_t}}, \\ \Sigma_{12} &= \Sigma_{21} = \sigma_{\theta_W\theta_R} - \frac{\sigma_{\theta\theta_W}\sigma_{\theta\theta_R}\sigma^2_{\theta_T} - \sigma_{\theta\theta_W}\sigma_{\theta_T\theta_R}\sigma_{\theta\theta_T} - \sigma_{\theta\theta_R}\sigma_{\theta_T\theta_W}\sigma_{\theta\theta_T} + \sigma_{\theta_T\theta_R}\sigma_{\theta_T\theta_W}\sigma^2_{\theta}}{\sigma^2_{\theta}\sigma^2_{\theta_t} - \sigma^2_{\theta\theta_t}}, \\ \Sigma_{22} &= \sigma^2_{\theta_R} - \frac{\sigma^2_{\theta\theta_R}\sigma^2_{\theta_T} - 2\sigma_{\theta\theta_R}\sigma_{\theta_T\theta_R}\sigma_{\theta\theta_T} + \sigma^2_{\theta_T\theta_R}\sigma^2_{\theta}}{\sigma^2_{\theta}\sigma^2_{\theta_t} - \sigma^2_{\theta\theta_t}}.\end{aligned}$$

Therefore, θ_{Wi} and θ_{Ri} are conditionally independent given θ_{Yi} if and only if $\mathrm{Cov}(\theta_{Wi}, \theta_{Ri}|\theta_{Yi}, \theta_{Ti}) = \Sigma_{12} = 0$. □

Proposition 3.1 leads to the following formulation:

$$\begin{aligned}\theta_{Ti} &= \eta_T\theta_{Yi} + \sqrt{1-(\eta_T)^2}\zeta_{Ti}, \\ \theta_{Wi} &= \eta_W \sin(\alpha_W)\theta_{Yi} + \eta_W \cos(\alpha_W)\zeta_{Ti} + \sqrt{1-(\eta_W)^2}\zeta_{Wi}, \\ \theta_{Ri} &= \eta_R \sin(\alpha_R)\theta_{Yi} + \eta_R \cos(\alpha_R)\zeta_{Ti} + \sqrt{1-(\eta_R)^2}\zeta_{Ri},\end{aligned} \tag{11}$$

where $\eta_T \in (-1,1)$, $\zeta_{Ti} \sim N(0,1)$, $\eta_W \in (-1,1)$, $\alpha_W \in (0,\pi]$, $\zeta_{Wi} \sim N(0,1)$, $\eta_R \in (-1,1)$, $\alpha_R \in (0,\pi]$, $\zeta_{Ri} \sim N(0,1)$, and ζ_{Ti}, ζ_{Wi},ζ_{Ri} are independent of θ_{Yi}, for $i = 1,\ldots,n$. It is worth noting that (η_W,α_W) and (η_R,α_R) can be viewed as polar coordinate systems.

In this formulation, $\boldsymbol{\mu} = (0,0,0,0)'$ and Σ are given by

$$\begin{bmatrix} 1 & \eta_T & \eta_W \sin(\alpha_W) & \eta_R \sin(\alpha_R) \\ \eta_T & 1 & \sigma_{\theta_T\theta_W} & \sigma_{\theta_T\theta_R} \\ \eta_W \sin(\alpha_W) & \sigma_{\theta_T\theta_W} & 1 & \eta_W\eta_R \cos(\alpha_W - \alpha_R) \\ \eta_R \sin(\alpha_R) & \sigma_{\theta_T\theta_R} & \eta_W\eta_R \cos(\alpha_W - \alpha_R) & 1 \end{bmatrix}, \tag{12}$$

where

$$\begin{aligned}\sigma_{\theta_T\theta_W} &= \eta_W\left(\eta_T \sin(\alpha_W) + \sqrt{1-\eta_T^2}\cos(\alpha_W)\right), \\ \sigma_{\theta_T\theta_R} &= \eta_R\left(\eta_T \sin(\alpha_R) + \sqrt{1-\eta_T^2}\cos(\alpha_R)\right).\end{aligned}$$

After some algebra, we can show that condition (10) in Proposition 3.1 is satisfied and consequently θ_{Wi} and θ_{Ri} are conditionally independent given θ_{Yi} and θ_{Ti} for any i.

Remark 3.1. To guarantee that the matrix in (12) is positive definite, we need to ensure that all the leading principal minors of the matrix are positive. After some algebra, we obtain the 1×1, 2×2, 3×3, and 4×4 leading principal minors: 1, $1-\eta_T^2$, $(1-\eta_T^2)(1-\eta_W^2)$, and $(1-\eta_T^2)(1-\eta_W^2)(1-\eta_R^2)$. All are positive since $\eta_T,\eta_W,\eta_R \in (-1,1)$. Thus, we can assign independent priors to these parameters, since there are no constraints between them. Compared with the original joint distribution in (8) and (9), which requires additional constraints on the parameters to ensure positive definiteness, the

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

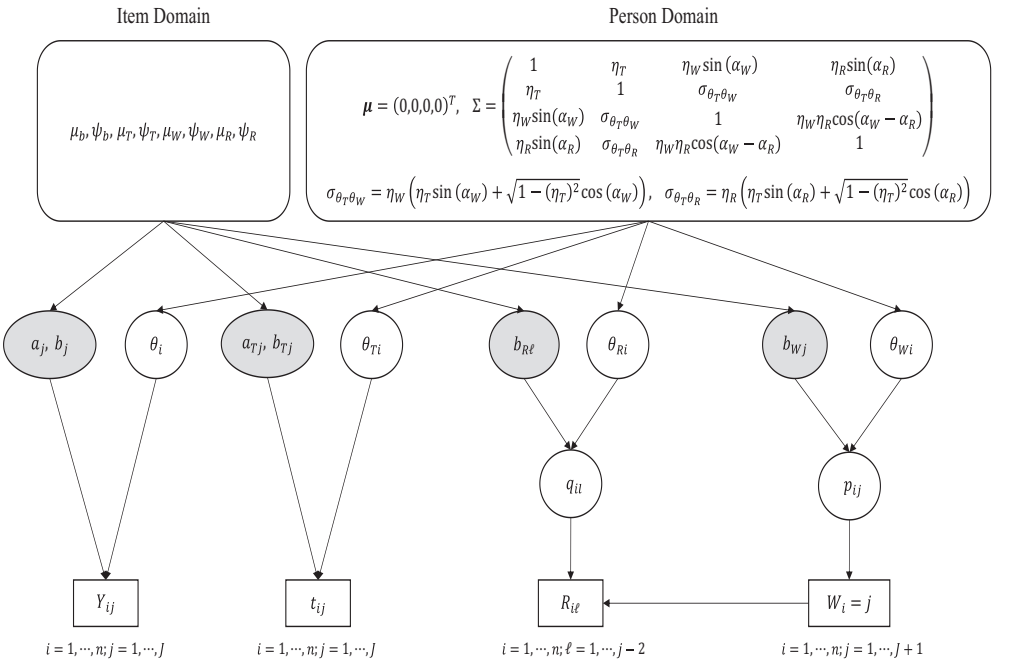


Figure 1. Diagram of the proposed model.

current formulation induced by conditional independence in (11) and (12) results in a more efficient implementation and better mixing of MCMC sampling.

Remark 3.2. Let $\mathbf{y}_{\text{obs}} = (\mathbf{y}'_{1,\text{obs}}, \ldots, \mathbf{y}'_{n,\text{obs}})'$ and $\mathbf{y}_{\text{mis}} = (\mathbf{y}'_{1,\text{mis}}, \ldots, \mathbf{y}'_{n,\text{mis}})'$, where $\mathbf{y}_{i,\text{obs}}$ and $\mathbf{y}_{i,\text{mis}}$ are the observed and missing responses for the i^{th} subject. Similarly, let $\mathbf{t}_{\text{obs}} = (\mathbf{t}'_{1,\text{obs}}, \ldots, \mathbf{t}'_{n,\text{obs}})'$ and $\mathbf{t}_{\text{mis}} = (\mathbf{t}'_{1,\text{mis}}, \ldots, \mathbf{t}'_{n,\text{mis}})'$, where $(\mathbf{t}_{i,\text{obs}}, \mathbf{t}_{i,\text{mis}})$ denote the response times for the observed and missing items, respectively, for the i^{th} subject, with the response times for the omitted items treated as missing data. Let $\Omega_{YT} = (\mathbf{a}_Y, \mathbf{b}_Y, \mathbf{a}_T, \mathbf{b}_T, \eta_T)$ and $\Omega_{WR} = (\mathbf{b}_W, \eta_W, \alpha_W, \mathbf{b}_R, \eta_R, \alpha_R)$ denote the collections of parameters for the response models and missing data models, respectively. Let V_L denote the collection of latent propensity variables, including $\theta_i, \zeta_{Ti}, \zeta_{Wi}, \zeta_{Ri}, i = 1, \ldots, n$.

As discussed in Ibrahim et al. (2005) and Wu et al. (2019), after integrating V_L, the missing data mechanism is ignorable if $P(\mathbf{W}, \mathbf{R}|\mathbf{y}, \mathbf{t}, \Omega_{yT}, \Omega_{WR}) = P(\mathbf{W}, \mathbf{R}|\Omega_{WR})$ for any $\mathbf{y}, \mathbf{t}, \Omega_{YT}$, and Ω_{WR}, that is, $\eta_W = \eta_R = 0$. After integrating V_L, the missing data mechanism is nonignorable if $P(\mathbf{W}, \mathbf{R}|\mathbf{y}, \mathbf{t}, \Omega_{YT}, \Omega_{WR}) = P(\mathbf{W}, \mathbf{R}|\mathbf{y}_{\text{obs}}, \mathbf{t}_{\text{obs}}, \Omega_{YT}, \Omega_{WR})$ for any $\mathbf{y}_{\text{mis}}, \mathbf{t}_{\text{mis}}$, Ω_{YT}, and Ω_{WR}, that is, $\eta_W \neq 0$ or $\eta_R \neq 0$.

Remark 3.3. The correlation between the latent not-reached propensity θ_{Wi} and the response propensity θ_{Yi} is measured by $\eta_W \sin(\alpha_W)$ and the correlation between θ_{Wi} and the latent propensity parameter for the response time θ_{Ti} is measured by $\eta_W(\eta_T \sin(\alpha_W) + \sqrt{1-\eta_T^2}\cos(\alpha_W))$. The correlations between θ_{Ri} and θ_{Yi}, θ_{Ri} and θ_{Ti} are similar, except that we replace η_W and α_W with η_R and α_R, respectively.

3.2. Difficulty parameters

We also assume a hierarchical framework for the difficulty parameters to improve the precision and convergence of the parameter estimation by borrowing information from other item parameters (Fox, 2010): $b_{Yj} \sim N(\mu_{bY}, \psi_Y^{-1}), b_{Tj} \sim N(\mu_T, {\psi_T}^{-1}), b_{Wj} \sim N(\mu_W, {\psi_W}^{-1})$, and $b_{Rj} \sim N(\mu_R, {\psi_R}^{-1}), j = 1, \ldots, J$.

A diagram of the proposed model within the hierarchical framework is given in Figure 1.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

4. Bayesian inference

Denote by $D_c = \{\mathbf{y}_i, \mathbf{t}_i, W_i, \mathbf{R}_i, i = 1, \ldots, n\}$ the set of complete data and $D_{\text{obs}} = \{\mathbf{y}_{i,\text{obs}}, \mathbf{t}_{i,\text{obs}}, W_i, \mathbf{R}_i, i = 1, \ldots, n\}$ the set of observed data for $\mathbf{y}, \mathbf{t}, \mathbf{W}, \mathbf{R}$. Let $D_{\mathbf{yWR},\text{obs}} = \{\mathbf{y}_{i,\text{obs}}, W_i, \mathbf{R}_i, i = 1, \ldots, n\}$ denote the set of observed data for $\mathbf{y}, \mathbf{W}, \mathbf{R}$, $D_{\mathbf{tWR},\text{obs}} = \{\mathbf{t}_{i,\text{obs}}, W_i, \mathbf{R}_i, i = 1, \ldots, n\}$ denote the set of observed data for $\mathbf{t}, \mathbf{W}, \mathbf{R}$, and $D_{\mathbf{WR},\text{obs}} = \{W_i, \mathbf{R}_i, i = 1, \ldots, n\}$ denote the set of data for $\mathbf{W}$ and $\mathbf{R}$, where there is no missingness.

4.1. Likelihood function

Let $\Omega = \{\mathbf{a}_Y, \mathbf{b}_Y, \mathbf{a}_T, \mathbf{b}_T, \eta_T, \sigma_T, \mathbf{b}_W, \eta_W, \alpha_W, \mathbf{b}_R, \eta_R, \alpha_R\}$ be the collection of all parameters, and also let $V_L = \{\theta_Y, \boldsymbol{\zeta}_T, \boldsymbol{\zeta}_W, \boldsymbol{\zeta}_R\}$ be the collection of all latent propensity variables, where $\theta_Y = (\theta_{Yi})$, $\boldsymbol{\zeta}_T = (\zeta_{Ti})$, $\boldsymbol{\zeta}_W = (\zeta_{Wi})$, and $\boldsymbol{\zeta}_R = (\zeta_{Ri})$. The complete data likelihood function is given by

$$\mathcal{L}(\Omega|D_c, V_L) = \prod_{i=1}^{n}\left[\prod_{j=1}^{w_i-1}\frac{\exp(y_{ij}a_{Yj}(\theta_{Yi}-b_{Yj}))}{1+\exp(a_{Yj}(\theta_{Yi}-b_{Yj}))}\frac{\sqrt{a_{Tj}}}{\sqrt{2\pi}t_{ij}}\exp\left\{-\frac{a_{Tj}(\log(t_{ij})-(b_{Tj}-\sigma_T\theta_{Ti}))^2}{2}\right\}\right]$$
$$\left\{\prod_{\ell=1}^{w_i-1}(1-p_{i\ell})\right\}^{\mathbf{1}(1<w_i\le J+1)} p_{iw_i}{}^{\mathbf{1}(w_i<J+1)}\left\{\prod_{\ell=1}^{w_i-2}q_{i\ell}^{r_{i\ell}}(1-q_{i\ell})^{1-r_{i\ell}}\right\}^{\mathbf{1}(w_i>2)}.$$

After integrating missing response measures $\mathbf{y}_{\text{mis}}$, the corresponding response times $\mathbf{t}_{\text{mis}}$, and all latent propensity variables V_L, the observed data likelihood function is given by

$$\mathcal{L}(\Omega|D_{\text{obs}}) = \prod_{i=1}^{n}\int\sum_{\mathbf{y}_{i,\text{mis}}}\left[\prod_{j=1}^{w_i-1}\frac{\exp(y_{ij}a_{Yj}(\theta_{Yi}-b_{Yj}))}{1+\exp(a_{Yj}(\theta_{Yi}-b_{Yj}))}\right.$$
$$\left.\frac{\sqrt{a_{Tj}}}{\sqrt{2\pi}t_{ij}}\exp\left\{-\frac{a_{Tj}(\log(t_{ij})-(b_{Tj}-\sigma_T(\eta_T\theta_{Yi}+\sqrt{1-(\eta_T)^2}\zeta_{Ti})))^2}{2}\right\}\right]d\mathbf{t}_{i,\text{mis}}$$
$$\left\{\prod_{\ell=1}^{w_i-1}(1-p_{i\ell})\right\}^{\mathbf{1}(1<w_i\le J+1)} p_{iw_i}{}^{\mathbf{1}(w_i<J+1)}\left\{\prod_{\ell=1}^{w_i-2}q_{i\ell}^{r_{i\ell}}(1-q_{i\ell})^{1-r_{i\ell}}\right\}^{\mathbf{1}(w_i>2)}$$
$$f(\theta_{Yi})f(\zeta_{Ti})f(\zeta_{Wi})f(\zeta_{Ri})d\theta_{Yi}d\zeta_{Ti}d\zeta_{Wi}d\zeta_{Ri},$$

where

$$p_{i\ell} = \frac{\exp(\eta_W\sin(\alpha_W)\theta_{Yi}+\eta_W\cos(\alpha_W)\zeta_{Ti}+\sqrt{1-\eta_W^2}\zeta_{Wi}-b_{W\ell})}{1+\exp(\eta_W\sin(\alpha_W)\theta_{Yi}+\eta_W\cos(\alpha_W)\zeta_{Ti}+\sqrt{1-\eta_W^2}\zeta_{Wi}-b_{W\ell})},$$
$$q_{i\ell} = \frac{\exp(\eta_R\sin(\alpha_R)\theta_{Yi}+\eta_R\cos(\alpha_R)\zeta_{Ti}+\sqrt{1-\eta_R^2}\zeta_{Ri}-b_{R\ell})}{1+\exp(\eta_R\sin(\alpha_R)\theta_{Yi}+\eta_R\cos(\alpha_R)\zeta_{Ti}+\sqrt{1-\eta_R^2}\zeta_{Ri}-b_{R\ell})}.$$

4.2. Joint prior and posterior distributions

The priors for the discrimination parameters a_{Yj} should be restricted to positive support (Fox, 2010). Common choices include log-normal distributions (Harwell & Baker, 1991; Mislevy, 1986; Patz & Junker, 1999). For the difficulty parameters, we assume hierarchical priors that improve the accuracy and convergence of parameter estimation via borrowing information from other item parameters (Fox, 2010). Priors with large variances and thus less informative are chosen for all other parameters (Fox, 2010; Gelman, 2006; Gelman & Hill, 2007). Specifically, we assume that the joint prior density can be

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

expressed as

$$\begin{aligned}\pi(\Omega) = {} & \pi(\mathbf{a}_Y)\pi(\mathbf{b}_Y)\pi(\mu_Y)\pi(\psi_Y)\\ & \pi(\eta_T)\pi(\mathbf{a}_T)\pi(\mathbf{b}_T)\pi(\sigma_T)\pi(\mu_T)\pi(\psi_T)\\ & \pi(\eta_W)\pi(\alpha_W)\pi(\mathbf{b}_W)\pi(\mu_W)\pi(\psi_W)\\ & \pi(\eta_R)\pi(\alpha_R)\pi(\mathbf{b}_R)\pi(\mu_R)\pi(\psi_R),\end{aligned}$$

with the following independent priors:

(i) $\{a_{Yj} \sim \text{log-normal}(0,1), j = 1,\ldots,J\}$, $\{b_{Yj} \sim N(\mu_Y, \psi_Y^{-1}), j = 1,\ldots,J\}$, $\mu_Y \sim N(0, 0.01^{-1}\psi_Y)$, $\psi_Y \sim \text{Gamma}(0.01, 0.01)$;
(ii) $\eta_T \sim U(-1,1)$, $\{a_{Tj} \sim \text{log-normal}(0,1), j = 1,\ldots,J\}$, $\{b_{Tj} \sim N(\mu_T, \psi_T{}^{-1}), j = 1,\ldots,J\}$, $\sigma_T \sim \text{log-normal}(0, 0.01^{-1})$, $\mu_T \sim N(0, 0.01^{-1}\psi_T)$, $\psi_T \sim \text{Gamma}(0.01, 0.01)$;
(iii) $\eta_W \sim U(-1,1)$, $\alpha_W \sim U(0,\pi]$, $\{b_{Wj} \sim N(\mu_W, \psi_W{}^{-1}), j = 1,\ldots,J\}$, $\mu_W \sim N(0, 0.01^{-1}\psi_W)$, $\psi_W \sim \text{Gamma}(0.01, 0.01)$;
(iv) $\eta_R \sim U(-1,1)$, $\alpha_R \sim U(0,\pi]$, $\{b_{Rj} \sim N(\mu_R, \psi_R{}^{-1}), j = 1,\ldots,J-1\}$, $\mu_R \sim N(0, 0.01^{-1}\psi_R)$, $\psi_R \sim \text{Gamma}(0.01, 0.01)$,

where $(\mu_Y, \mu_T, \mu_W, \mu_R)'$ and $(\psi_Y, \psi_T, \psi_W, \psi_R)'$ are the means and precisions of $(\mathbf{b}_Y', \mathbf{b}_T', \mathbf{b}_W', \mathbf{b}_R')'$. The joint posterior based on the observed data D_{obs} is written as

$$\pi(\Omega|D_{\text{obs}}) \propto \mathcal{L}(\Omega|D_{\text{obs}})\pi(\Omega). \tag{13}$$

4.3. Bayesian model comparison—LPML decomposition

Similarly to the conventional LPML (Ibrahim, Chen, & Sinha, 2001), we use the joint probability, $\prod_{i=1}^{n} f(\mathbf{y}_i, \mathbf{t}_i, W_i, \mathbf{R}_i|\Omega)$, to measure the model's predictive ability using a leave-one-out idea. Let $D_{\text{obs}}^{(-i)} = \{\mathbf{y}_{\ell,\text{obs}}, \mathbf{t}_{\ell,\text{obs}}, W_\ell, \mathbf{R}_\ell, \ell = 1,\ldots,i-1,i+1,\ldots,n\}$ denote the observed data with $(\mathbf{y}_i, \mathbf{t}_i, W_i, \mathbf{R}_i)$ deleted. Let $D_{\mathbf{WR},\text{obs}}^{(-i)} = \{W_\ell, \mathbf{R}_\ell, \ell = 1,\ldots,i-1,i+1,\ldots,n\}$ denote the observed data with $(W_i, \mathbf{R}_i)$ deleted. Let Ω_1 and Ω_2 denote the parameters of $(\mathbf{y}, \mathbf{t})$ and $(\mathbf{W}, \mathbf{R})$, respectively. Specifically, $\Omega_1 = \{\mathbf{a}_Y, \mathbf{b}_Y, \mu_Y, \psi_Y, \mathbf{a}_T, \mathbf{b}_T, \mu_T, \psi_T, \eta_T, \sigma_T\}$ and $\Omega_2 = \{\mathbf{b}_W, \mu_W, \psi_W, \eta_W, \alpha_W, \mathbf{b}_R, \mu_R, \psi_R, \eta_R, \alpha_R\}$.

The conditional predictive ordinate (CPO) identity (Geisser & Eddy, 1979) for the i^{th} subject is defined as

$$\text{CPO}_i = \int f(\mathbf{y}_i, \mathbf{t}_i, W_i, \mathbf{R}_i|\Omega)\pi(\Omega|D_{\text{obs}}^{(-i)})d\Omega,$$

where $\pi(\Omega|D_{\text{obs}}^{(-i)}) = \frac{\prod_{j\neq i} f(\mathbf{y}_i, \mathbf{t}_i, W_i, \mathbf{R}_i|\Omega)\pi(\Omega)}{\int \prod_{j\neq i} f(\mathbf{y}_i, \mathbf{t}_i, W_i, \mathbf{R}_i|\Omega)\pi(\Omega)d\Omega}$ and LPML is defined as $\text{LPML} = \sum_{i=1}^{n} \log \text{CPO}_i$,

Using CPO identity III (Zhang et al., 2017), we have

$$\text{CPO}_i = \frac{f(\mathbf{y}_i, \mathbf{t}_i, W_i, \mathbf{R}_i|\Omega)\pi(\Omega|D_{\text{obs}}^{(-i)})}{\pi(\Omega|D_{\text{obs}})}, \tag{14}$$

for any value of Ω. CPO $_i$ can be decomposed as

$$\text{CPO}_i = \text{CPO}_{i,\mathbf{yt}} \cdot \text{CPO}_{i,\mathbf{WR}|\mathbf{yt}},$$

where

$$\text{CPO}_{i,\mathbf{yt}} = \frac{f(\mathbf{y}_i, \mathbf{t}_i|\Omega_1)\pi(\Omega_1|D_{\text{obs}}^{(-i)})}{\pi(\Omega_1|D_{\text{obs}})} \tag{15}$$

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

and

$$\text{CPO}_{i,\mathbf{WR}|\mathbf{yt}} = \frac{f(W_i,\mathbf{R}_i|\mathbf{y}_i,\mathbf{t}_i,\Omega)\pi(\Omega_2|\Omega_1,D_{\text{obs}}^{(-i)})}{\pi(\Omega_2|\Omega_1,D_{\text{obs}})}. \tag{16}$$

Proposition 4.1. *The total LPML under the proposed joint model can be decomposed as*

$$LPML = LPML_{\mathbf{yt}} + LPML_{\mathbf{WR}|\mathbf{yt}},$$

where $LPML_{\mathbf{yt}} = \sum_{i=1}^{n} \log CPO_{i,\mathbf{yt}}$ *and* $LPML_{\mathbf{WR}|\mathbf{yt}} = \sum_{i=1}^{n} \log CPO_{i,\mathbf{WR}|\mathbf{yt}}$.

To compute $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$, we start with Remark 4 of Zhang et al. (2017),

$$\frac{\pi(\bar{\Omega}_1|D_{\text{obs}}^{(-i)})}{\pi(\bar{\Omega}_1|D_{\text{obs}})} = \text{CPO}_i \int \frac{\pi(\Omega_2|\bar{\Omega}_1,D_{\text{obs}})}{f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1)f(\mathbf{t}_i|\bar{\Omega}_1)f(W_i,\mathbf{R}_i|\mathbf{y}_i,\mathbf{t}_i,\bar{\Omega}_1,\Omega_2)} d\Omega_2, \tag{17}$$

where we plug in a fixed value of Ω_1 such as the posterior mean, denoted by $\bar{\Omega}_1$.

Together with (15), we obtain

$$\begin{aligned}
&\left[\text{CPO}_{i,\mathbf{WR}|\mathbf{yt}}\right]^{-1} \\
&= f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1)f(\mathbf{t}_i|\bar{\Omega}_1) \int \frac{\pi(\Omega_2|\bar{\Omega}_1,D_{\text{obs}})}{f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1)f(\mathbf{t}_i|\bar{\Omega}_1)f(W_i,\mathbf{R}_i|\mathbf{y}_i,\mathbf{t}_i,\bar{\Omega}_1,\Omega_2)} d\Omega_2 \\
&= f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1) \int \frac{k(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,\Omega_2,\mathbf{y}_i,\mathbf{t}_i,W_i,\mathbf{R}_i)\pi(\Omega_2,\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,D_{\text{obs}})}{f(\mathbf{y}_i|\theta_{Yi},\bar{\Omega}_1)f(W_i,\mathbf{R}_i|\theta_{Wi},\theta_{Ri},\Omega_2)f(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\mathbf{t}_i)} \\
&\quad d\boldsymbol{\theta}_Y d\theta_W d\theta_R d\Omega_2,
\end{aligned} \tag{18}$$

where $\int k(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,\Omega_2,\mathbf{y}_i,\mathbf{t}_i,W_i,\mathbf{R}_i)d\theta_{Yi}d\theta_{Wi}d\theta_{Ri} = 1$ and $\pi(\Omega_2,\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,D_{\text{obs}})$ is the conditional posterior density of $(\Omega_2,\theta_{Yi},\theta_{Wi},\theta_{Ri})$ given $\bar{\Omega}_1$ and D_{obs}. In both simulations and real data analysis, we choose a multivariate normal distribution for $k(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,\Omega_2,\mathbf{y}_i,\mathbf{t}_i,W_i,\mathbf{R}_i)$ that approximates $f(\mathbf{y}_i|\theta_{Yi},\bar{\Omega}_1)f(W_i,\mathbf{R}_i|\theta_{Wi},\theta_{Ri},\Omega_2)f(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\mathbf{t}_i)$. The remaining two conditional densities that need to be computed are $f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1)$ and $f(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\mathbf{t}_i)$.

To compute $f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1)$, we apply the formula:

$$f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1) = \int \left\{\prod_{j=1}^{w_i-1} f(y_{ij}|\theta_{Yi},\bar{\Omega}_1)^{1-r_{ij}}\right\} f(\theta_{Yi}|\mathbf{t}_i,\bar{\Omega}_1)d\theta_{Yi},$$

where

$$f(\theta_{Yi}|\mathbf{t}_i,\bar{\Omega}_1) \propto f(\theta_{Yi}) \int \left\{\prod_{j=1}^{w_i-1} f(t_{ij}|\theta_{Ti},\bar{\Omega}_1)^{1-r_{ij}}\right\} f(\theta_{Ti}|\theta_{Yi})d\theta_{Ti}$$

and $\theta_{Ti}|\theta_{Yi} \sim N(\sigma_{\theta\theta_T}\theta_{Yi}, 1-\sigma^2_{\theta\theta_T})$. After some algebra, we obtain

$$\theta_{Yi}|\mathbf{t}_i,\bar{\Omega}_1 \sim N\left(\frac{-\tilde{\eta}_T\tilde{\sigma}_T\sum_{j=1}^{w_i-1}(1-r_{ij})\tilde{a}_{Tj}(\log(t_{ij})-\bar{b}_{Tj})}{\tilde{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\tilde{a}_{Tj}+1}, \frac{(1-\tilde{\eta}_T^2)\tilde{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\tilde{a}_{Tj}+1}{\tilde{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\tilde{a}_{Tj}+1}\right).$$

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Thus,

$$f(\mathbf{y}_i|\mathbf{t}_i,\bar{\Omega}_1) = \int \left[\prod_{j=1}^{w_i-1}\left\{\frac{\exp(y_{ij}\bar{a}_{Yj}(\theta_{Yi}-\bar{b}_{Yj}))}{1+\exp(\bar{a}_{Yj}(\theta_{Yi}-\bar{b}_{Yj}))}\right\}^{1-r_{ij}}\right]\frac{\sqrt{\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Tj}+1}}{\sqrt{2\pi\left\{(1-\bar{\eta}_T^2)\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Tj}+1\right\}}}$$

$$\exp\left\{-\frac{\left\{(\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Tj}+1)\theta_{Yi}+\bar{\eta}_T\bar{\sigma}_T\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Tj}(\log(t_{ij})-\bar{b}_{Tj})\right\}^2}{2\left\{(1-\bar{\eta}_T^2)\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Tj}+1\right\}(\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Tj}+1)}\right\}d\theta_{Yi}.$$

To determine $f(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\mathbf{t}_i)$, we apply the formula:

$$f(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\mathbf{t}_i) \propto f(\theta_{Yi},\theta_{Wi},\theta_{Ri})\int\left\{\prod_{j=1}^{w_i-1}f(t_{ij}|\theta_{Ti},\bar{\Omega}_1)\right\}f(\theta_{Ti}|\theta_{Yi},\theta_{Wi},\theta_{Ri})d\theta_{Ti},$$

where $\theta_{Ti}|\theta_{Yi},\theta_{Wi},\theta_{Ri} \sim N(\gamma_1\theta_i+\gamma_2\theta_{Wi}+\gamma_3\theta_{Ri},\gamma_4)$, $(\gamma_1,\gamma_2,\gamma_3)' = \Sigma'_{2,134}\Sigma_{134}^{-1}$, $\gamma_4 = 1-\Sigma'_{2,134}\Sigma_{134}^{-1}\Sigma_{2,134}$, $\Sigma_{2,134} = (\sigma_{\theta\theta_T},\sigma_{\theta_T\theta_W},\sigma_{\theta_T\theta_R})'$, and Σ_{134} is the covariance matrix corresponding to θ_{Yi}, θ_{Wi}, and θ_{Ri}. After some algebra, we have

$$(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\mathbf{t}_i)' \sim N(\boldsymbol{\mu}_i^*,\Sigma_i^*),$$

where $\boldsymbol{\mu}_i^* = -\frac{\bar{\sigma}_T\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Yj}(\log(t_{ij})-\bar{b}_{Tj})}{\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}\bar{a}_{Yj}\gamma_4+1}\Sigma_i^*\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\end{bmatrix}$,

$$\Sigma_i^{*-1} = \begin{bmatrix}\frac{1-\sigma^2_{\theta_W\theta_R}}{\gamma_0}+\kappa_{i11} & \frac{\sigma_{\theta\theta_R}\sigma_{\theta_W\theta_R}-\sigma_{\theta\theta_W}}{\gamma_0}+\kappa_{i12} & \frac{\sigma_{\theta\theta_W}\sigma_{\theta_W\theta_R}-\sigma_{\theta\theta_R}}{\gamma_0}+\kappa_{i13}\\ \frac{\sigma_{\theta\theta_R}\sigma_{\theta_W\theta_R}-\sigma_{\theta\theta_W}}{\gamma_0}+\kappa_{i12} & \frac{1-\sigma^2_{\theta\theta_R}}{\gamma_0}+\kappa_{i22} & \frac{\sigma_{\theta\theta_W}\sigma_{\theta\theta_R}-\sigma_{\theta_W\theta_R}}{\gamma_0}+\kappa_{i23}\\ \frac{\sigma_{\theta\theta_W}\sigma_{\theta_W\theta_R}-\sigma_{\theta\theta_R}}{\gamma_0}+\kappa_{i13} & \frac{\sigma_{\theta\theta_W}\sigma_{\theta\theta_R}-\sigma_{\theta_W\theta_R}}{\gamma_0}+\kappa_{i23} & \frac{1-\sigma^2_{\theta\theta_W}}{\gamma_0}+\kappa_{i33}\end{bmatrix},$$

$\kappa_{ikk'} = \frac{\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}(1-r_{ij})\bar{a}_{Yj}\gamma_k\gamma_{k'}}{\bar{\sigma}_T^2\sum_{j=1}^{w_i-1}\bar{a}_{Yj}\gamma_4+1}$, and $\gamma_0 = 1-\sigma^2_{\theta\theta_W}-\sigma^2_{\theta\theta_R}-\sigma^2_{\theta_W\theta_R}+2\sigma_{\theta\theta_W}\sigma_{\theta\theta_R}\sigma_{\theta_W\theta_R}$.

Similarly to a traditional LPML, a higher value of $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ indicates a more favorable model in terms of prediction.

Remark 4.1. To measure the extent to which the missing data mechanism is nonignorable, we introduce $\Delta\text{LPML} = \text{LPML}_{\mathbf{WR}|\mathbf{yt}} - \text{LPML}_{\mathbf{WR}}$, where $\text{LPML}_{\mathbf{WR}} = \sum_{i=1}^n \log \text{CPO}_{i,\mathbf{WR}}$ and $\text{CPO}_{i,\mathbf{WR}}$ is calculated solely based on the joint model of $\mathbf{W}$ and $\mathbf{R}$. Specifically, $\left[\text{CPO}_{i,\mathbf{WR}}\right]^{-1} = \int \frac{k(\theta_{Wi},\theta_{Ri}|\Omega_2,W_i,\mathbf{R}_i)\pi(\Omega_2,\theta_{Wi},\theta_{Ri}|D_{\text{obs}})}{f(W_i,\mathbf{R}_i|\theta_{Wi},\theta_{Ri},\Omega_2)f(\theta_{Wi},\theta_{Ri})}$ $d\theta_W d\theta_R d\Omega_2$, where $k(\theta_{Wi},\theta_{Ri}|\Omega_2,W_i,\mathbf{R}_i)$ is a bivariate normal distribution that approximates the joint posterior distribution of Ω_2.

$\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ is reduced to $\text{LPML}_{\mathbf{WR}}$ and consequently ΔLPML is 0 if the missing data mechanism is ignorable, that is, $\eta_W = \eta_R = 0$. This formulation, by integrating the latent variables associated with $\mathbf{t}$, enables the evaluation of multidimensional integrals, making it computationally efficient. Furthermore, it simplifies the comparison between the joint model of $(\mathbf{y},\mathbf{t})$ and the single outcome model of $\mathbf{y}$, as $\text{CPO}_{i,\mathbf{WR}|\mathbf{yt}}$ and $\text{CPO}_{i,\mathbf{WR}|\mathbf{y}}$ share the same dimension.

Remark 4.2. To assess the importance of modeling item response ($\mathbf{y}$) and response time ($\mathbf{t}$) jointly versus individually ($\mathbf{y}$–only or $\mathbf{t}$–only) in identifying the missing data mechanism, we consider the following single response models. The first single response model we compare is the one that includes

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

only **y**, where (11) is reformulated as follows:

$$\theta_{Wi} = \eta_W \theta_{Yi} + \sqrt{1-\eta_W^2}\zeta_{Wi},$$
$$\theta_{Ri} = \eta_R \theta_{Yi} + \sqrt{1-\eta_R^2}\zeta_{Ri}. \tag{19}$$

Consequently,

$$\left[\text{CPO}_{i,\mathbf{WR}|\mathbf{y}}\right]^{-1} = f(\mathbf{y}_i|\bar{\Omega}_1) \int \frac{k(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,\Omega_2,\mathbf{y}_i,W_i,\mathbf{R}_i)\pi(\Omega_2,\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,D_{\mathbf{yWR},\text{obs}})}{f(\mathbf{y}_i|\theta_{Yi},\bar{\Omega}_1)f(W_i,\mathbf{R}_i|\theta_{Wi},\theta_{Ri},\Omega_2)f(\theta_{Yi},\theta_{Wi},\theta_{Ri})} d\boldsymbol{\theta}_i d\theta_W d\theta_R d\Omega_2,$$

where $\int k(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,\Omega_2,\mathbf{y}_i,W_i,\mathbf{R}_i)d\theta_{Yi}d\theta_{Wi}d\theta_{Ri} = 1$. The second single response model we compare is the one that includes only **t**, where (11) is reformulated as follows:

$$\theta_{Wi} = \eta_W \theta_{Ti} + \sqrt{1-\eta_W^2}\zeta_{Wi},$$
$$\theta_{Ri} = \eta_R \theta_{Ti} + \sqrt{1-\eta_R^2}\zeta_{Ri}. \tag{20}$$

Consequently,

$$\left[\text{CPO}_{i,\mathbf{WR}|\mathbf{t}}\right]^{-1} = f(\mathbf{t}_i|\bar{\Omega}_1) \int \frac{k(\theta_{Ti},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,\Omega_2,\mathbf{t}_i,W_i,\mathbf{R}_i)\pi(\Omega_2,\theta_{Ti},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,D_{\mathbf{tWR},\text{obs}})}{f(\mathbf{t}_i|\theta_{Ti},\bar{\Omega}_1)f(W_i,\mathbf{R}_i|\theta_{Wi},\theta_{Ri},\Omega_2)f(\theta_{Ti},\theta_{Wi},\theta_{Ri})} d\boldsymbol{\theta}_{Ti} d\theta_W d\theta_R d\Omega_2,$$

where $\int k(\theta_{Yi},\theta_{Wi},\theta_{Ri}|\bar{\Omega}_1,\Omega_2,\mathbf{t}_i,W_i,\mathbf{R}_i)d\theta_{Ti}d\theta_{Wi}d\theta_{Ri} = 1$.

We then compute $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ - $\text{LPML}_{\mathbf{WR}|\mathbf{y}}$ and $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ - $\text{LPML}_{\mathbf{WR}|\mathbf{t}}$, respectively. A positive value indicates that the joint model provides a better fit for the missing data model. We can also compare between $\text{LPML}_{\mathbf{WR}|\mathbf{y}}$ and $\text{LPML}_{\mathbf{WR}|\mathbf{t}}$, where a larger value indicates that the corresponding outcome model provides better identification of the missing data mechanism.

5. Simulations

In this section, we conduct extensive simulation studies to examine the performance of $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ on model selection and the performance of the proposed nonignorable and ignorable missing data mechanisms in modeling the simulated missing data. In all Bayesian computations, we use 5,000 MCMC samples, which are taken from every tenth iteration, after a burn-in of 5,000 iterations for each model to compute all the posterior estimates, including posterior means, posterior standard deviations, 95% highest posterior density (HPD) intervals, and $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$. The MCMC sampling code is written with R package *nimble* (de Valpine et al., 2017) with double-precision accuracy. The code for computing the $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ is written in FORTRAN 95 using IMSL subroutines with double-precision accuracy. The convergence of the MCMC samples is checked by the R package "mcmcplots" using R version 4.5.2. Convergence is reached approximately after 5,000 iterations. The codes for nimble and Fortran are provided in the Supplementary Material.

5.1. Simulation Scenarios I and II

The purpose of simulations in this section is to investigate the impact of misspecification of the missing data model under a setting similar to that of the real data. In Simulation Scenarios I and II, we first generate $S = 1{,}000$ replications. For each replication, we set the total number of subjects (n) equal to 2,000 and the total number of items (J) equal to 12 to mimic the real data. The binary responses y_{ij}

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

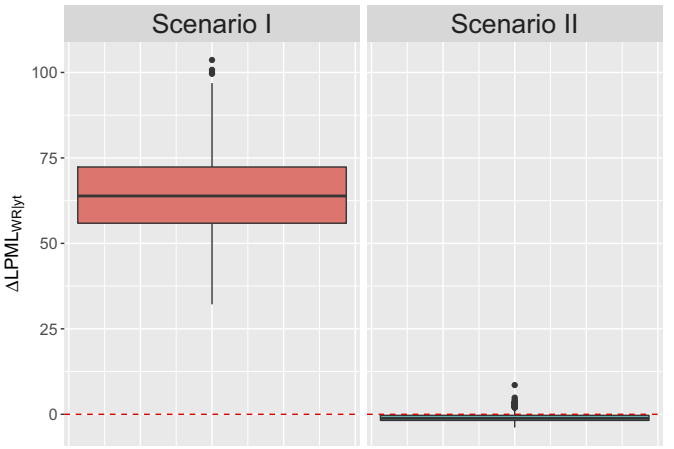


Figure 2. Boxplots of $\Delta\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) when the true missing data mechanisms are nonignorable (Simulation Scenario I in red) and ignorable (Simulation Scenario II in cyan).

are generated based on the item response model in (1), where the a_{Yj}s are sampled from a truncated normal distribution with a lower bound of 0.2 and an upper bound of 1, $b_{Yj} \sim N(\mu_Y, \psi_Y^{-1})$ with $\mu_Y = 1$ and $\psi_Y = 1$ for $j = 1, \ldots, J$. The response times t_{ij} are generated based on the model in (2), where a_{Tj}s are sampled from a truncated normal distribution with a lower bound of 0.5 and an upper bound of 4, $\eta_T = -0.4$, $b_{Tj} \sim N(\mu_T, \psi_T^{-1})$ with $\mu_T = 2$ and $\psi_T = 1$ for $j = 1, \ldots, J$. For the missing data models in Simulation Scenarios I and II, the not-reached items W_i are generated based on models in (3)–(5), where $b_{Wj} \sim N(\mu_W, \psi_W^{-1})$ with $\mu_W = 6$ and $\psi_W = 1$ for $j = 1, \ldots, J$, to mimic the real data. The omitted items, $R_{i\ell}$s, are generated based on models in (6) and (7), where $b_{Rj} \sim N(\mu_R, \psi_R^{-1})$ with $\mu_R = 2$ and $\psi_R = 1$ for $j = 1, \ldots, J$. To investigate the associations between the latent propensity parameters, we independently generate $\theta_{Yi}, \zeta_{Ti}, \zeta_{Wi}, \zeta_{Ri} \sim N(0,1)$ for each replication, $i = 1, \ldots, n$.

The only differences between Simulation Scenarios I and II are the choices of η_W, α_W, η_R, and α_R. In Simulation Scenario I, we set $\eta_W = -0.3$, $\alpha_W = 1.3$, $\eta_R = -0.5$, and $\alpha_R = 1.7$ to mimic the real data in Section 6. In Simulation Scenario II, all values are set to 0 to generate an ignorable missing data mechanism. In both settings, the missing data percentages of the simulated datasets are around 20%, with 15% due to omission behavior and 5% due to not-reached behavior, which is similar to the real data. The missing percentages by item are provided in Figures S1 and S2 in the Supplementary Material.

We fit both nonignorable and ignorable models to the simulated nonignorable datasets in Simulation Scenario I and the simulated ignorable datasets in Simulation Scenario II. The boxplots of the differences values $\Delta\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) under these two models are given in Figure 2. For Simulation Scenario I where the true missing data mechanism is nonignorable, the median of the values of $\text{LPML}_{\mathbf{WR|yt}}$ under the nonignorable model (−9674.02) is greater than that under the ignorable model (−9738.74). Furthermore, the entire boxplot (Median=63.86, IQR=(55.90, 72.34)) for $\Delta\text{LPML}_{\mathbf{WR|yt}}$ is above 0 (Figure 2 left panel in red), indicating that $\text{LPML}_{\mathbf{WR|yt}}$ correctly selects the best model in this case. For Simulation Scenario II where the true missing data mechanism is ignorable, the median of the values of $\text{LPML}_{\mathbf{WR|yt}}$ under the ignorable model (−9240.21) is slightly greater than that under the nonignorable model (−9241.49). The horizontal $y = 0$ line intersects the whisker of the boxplot (Median= −1.18, IQR=(−1.79, −0.29)). This is expected as the nonignorable model is approximately ignorable when the estimates of η_T, η_W, and η_R are close to zero, as observed in Scenario II. However, the $\Delta\text{LPML}_{\mathbf{WR|yt}}$ measure accounts for the dimension penalty, as reflected in the boxplot. The entire box for $\Delta\text{LPML}_{\mathbf{WR|yt}}$ is below 0 (Figure 2 right panel in cyan), indicating that $\text{LPML}_{\mathbf{WR|yt}}$ effectively selects the true model in this case. In conclusion, $\text{LPML}_{\mathbf{WR|yt}}$ can serve as a good criterion for selecting a better model in these two different settings.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

In Tables 1 and 2, we report the differences between the posterior mean and the true value (Bias), the standard deviation of the estimate (SD), the root mean squared error of the posterior mean (RMSE), and the coverage probability (CP) of the 95% HPD interval for each parameter in the item response, response time, and missing data models under the optimal models in Simulations I (i.e., nonignorable) and II (i.e., ignorable), respectively. In both tables, the Bias and SDs of the posterior estimates are relatively small, SDs are similar to RMSEs, and most of the CPs are around the 95% nominal level. The CP for b_{Y7} in Simulation Scenario I, which corresponds to the item with the smallest value of $\mathbf{b}_Y$ and the highest missing percentage (median=58.50%), is slightly lower than the nominal level. Similarly, the CP for b_{Y4} and b_{W11} in Simulation Scenario II, which correspond to the item with the largest value of $\mathbf{b}_Y$ and the highest missing percentage (median=57.45%), and the item with the largest value of $\mathbf{b}_W$ and the second highest missing percentage (median=48.20%), respectively, are slightly lower than the nominal level. These CPs are improved when sample size increases, as shown in Simulation Scenario VIII.

To further investigate the performance of the θ estimates, we conduct an additional simulation under the same setting as Simulation Scenario I, except that θ is fixed across repetitions. The posterior summaries of θ also exhibit good performance, with the following median (IQR) values: bias= −0.058 (−0.363, −0.021), SD=0.701 (0.688, 0.704), RMSE=0.606 (0.523, 0.681), and CP=0.976 (0.940, 0.989).

Under the same setting of Simulation Scenario I, we compare our methods with multivariate imputation by chained equations (MICE) methods (Rubin, 1987; Van Buuren & Groothuis-Oudshoorn, 2011) in Table 3. Twenty mutiple imputations are performed. The same item response and response time models in (1) and (2) and the same hierarchical framework and reparameterization techniques in Section 3 are assumed for fair comparison. Weighted predicted mean matching method is used for the imputation step. The posterior summaries of the response models are good, while poor CPs are observed for the parameters in the response time model. The two missing data behaviors are not studied under MICE and the method is computationally intensive.

We also compare our methods with the RTprocess model (Lu & Wang, 2020) in Table 4, where the item omission behavior is modeled as a type of censoring based on the response time model. Most of the posterior summaries of the response models are good, while large biases and poor CPs are observed for the parameters in the response time model and the correlation parameter η_T. This is expected as the method assumes the omission and not-reached behaviors are solely due to time limitations. However, it does not account for the propensities of omission and not-reach that are incorporated in the simulation design. Additionally, we compare these three methods when data are generated from the RTprocess model, with a similar percentage of missingness (15% due to omission behavior and 5% due to not-reached behavior). Posterior summaries under the RTprocess model (Table S1 in the Supplementary Material) are good, as expected, since this is the data-generating model. Under the proposed nonignorable model, posterior summaries of most parameters (Table S2 in the Supplementary Material) are also satisfactory, except for several item-specific parameters in the response time model that correspond to high missingness, as well as σ_T and η_T. Results under MICE are the worst under this setting (see Table S3 in the Supplementary Material).

5.2. Simulation Scenarios III and IV

The purpose of simulations in this section is to investigate the impact of misspecification of the missing data model when the missing data percentages (30%) are higher than those in Scenarios I and II, with 20% due to omission behavior and 10% due to not-reached behavior. In Simulation Scenarios III and IV, we set $\mu_W = 5.5, \mu_R = 1.5$ to increase the missing data percentage while keeping all other conditions the same as in Scenarios I and II.

Similarly, we fit both nonignorable and ignorable models to the simulated nonignorable datasets in Simulation Scenario III and the simulated ignorable datasets in Simulation Scenario IV. The boxplots of the differences values $\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) under these two models are given in Figure 3. For Simulation Scenario III where the true missing data mechanism is nonignorable, the median of the values of $\text{LPML}_{\mathbf{WR|yt}}$ under the nonignorable model (−11042.51) is greater than that

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 1. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the nonignorable missing data mechanism in Simulation Scenario I ($n = 2{,}000$, $J = 12$, median sizes of ηs, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.012	0.092	0.096	0.942	b_{Y1}	0.026	0.114	0.119	0.930
a_{Y2}	0.003	0.074	0.073	0.948	b_{Y2}	0.008	0.131	0.124	0.950
a_{Y3}	−0.001	0.071	0.068	0.951	b_{Y3}	0.053	0.237	0.223	0.955
a_{Y4}	0.026	0.105	0.110	0.937	b_{Y4}	−0.009	0.147	0.147	0.944
a_{Y5}	0.007	0.089	0.089	0.941	b_{Y5}	0.009	0.153	0.148	0.954
a_{Y6}	0.015	0.067	0.062	0.966	b_{Y6}	−0.012	0.218	0.191	0.960
a_{Y7}	0.052	0.083	0.087	0.953	b_{Y7}	0.271	0.322	0.368	0.887
a_{Y8}	−0.008	0.080	0.079	0.952	b_{Y8}	0.012	0.121	0.118	0.954
a_{Y9}	0.008	0.118	0.111	0.955	b_{Y9}	0.008	0.137	0.128	0.959
a_{Y10}	−0.005	0.103	0.107	0.936	b_{Y10}	0.011	0.082	0.082	0.948
a_{Y11}	−0.005	0.090	0.090	0.943	b_{Y11}	0.020	0.120	0.119	0.951
a_{Y12}	−0.004	0.085	0.084	0.943	b_{Y12}	0.001	0.108	0.100	0.957
a_{T1}	0.010	0.104	0.109	0.933	b_{T1}	0.001	0.019	0.019	0.952
a_{T2}	0.001	0.087	0.086	0.952	b_{T2}	0.000	0.018	0.017	0.942
a_{T3}	0.000	0.088	0.088	0.940	b_{T3}	0.000	0.020	0.019	0.949
a_{T4}	0.002	0.057	0.055	0.951	b_{T4}	−0.001	0.021	0.021	0.954
a_{T5}	0.003	0.138	0.132	0.953	b_{T5}	−0.001	0.017	0.017	0.934
a_{T6}	0.001	0.088	0.087	0.949	b_{T6}	−0.001	0.018	0.017	0.956
a_{T7}	0.006	0.079	0.079	0.941	b_{T7}	0.001	0.030	0.030	0.952
a_{T8}	0.001	0.072	0.072	0.941	b_{T8}	0.000	0.020	0.021	0.948
a_{T9}	−0.001	0.115	0.116	0.948	b_{T9}	0.000	0.019	0.019	0.960
a_{T10}	0.004	0.096	0.097	0.940	b_{T10}	0.000	0.019	0.018	0.960
a_{T11}	−0.001	0.066	0.064	0.954	b_{T11}	0.000	0.021	0.021	0.960
a_{T12}	−0.001	0.100	0.101	0.947	b_{T12}	0.000	0.019	0.019	0.946
b_{W1}	−0.011	0.541	0.516	0.959	b_{R1}	0.004	0.059	0.059	0.943
b_{W2}	−0.137	0.614	0.543	0.948	b_{R2}	−0.002	0.093	0.093	0.938
b_{W3}	0.048	0.142	0.148	0.950	b_{R3}	0.003	0.066	0.065	0.946
b_{W4}	0.036	0.470	0.448	0.962	b_{R4}	0.005	0.094	0.093	0.953
b_{W5}	0.086	0.352	0.365	0.960	b_{R5}	0.005	0.071	0.068	0.956
b_{W6}	0.099	0.367	0.392	0.954	b_{R6}	−0.014	0.151	0.145	0.949
b_{W7}	0.037	0.130	0.135	0.945	b_{R7}	0.007	0.057	0.056	0.947
b_{W8}	0.072	0.230	0.251	0.937	b_{R8}	0.005	0.081	0.081	0.952
b_{W9}	0.087	0.257	0.280	0.940	b_{R9}	0.006	0.062	0.062	0.940
b_{W10}	0.097	0.465	0.486	0.954	b_{R10}	0.009	0.081	0.081	0.946
b_{W11}	0.090	0.351	0.380	0.942	b_{R11}	0.007	0.098	0.097	0.947

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 1. Continued

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
b_{W12}	0.006	0.076	0.075	0.955	σ_T	0.000	0.008	0.008	0.940
η_T	0.003	0.036	0.038	0.941					
η_W	0.024	0.105	0.107	0.940					
α_W	0.074	0.385	0.264	0.966					
η_R	0.002	0.045	0.045	0.947					
α_R	0.010	0.091	0.090	0.952					

under the ignorable model (−11106.94). Additionally, the box (Median=66.46, IQR=(58.64, 74.11)) for the $\Delta\text{LPML}_{\mathbf{WR|yt}}$ is above 0 (Figure 3 left panel in red), indicating that the nonignorable model outperforms the ignorable model when the true missing data mechanism is nonignorable. For Simulation Scenario IV where the true missing data mechanism is ignorable, the median of the values of $\text{LPML}_{\mathbf{WR|yt}}$ under the ignorable model (−10452.53) is slightly greater than that under the nonignorable model (−10453.09). Posterior estimates of η_T, η_W, and η_R under the nonignorable model are close to 0, making it similar to the ignorable model. A key advantage of $\text{LPML}_{\mathbf{WR|yt}}$ is that it accounts for dimension penalty. The entire box (Median= −1.16, IQR= (−1.87, −0.31)) for the $\Delta\text{LPML}_{\mathbf{WR|yt}}$ is below 0 (Figure 3 right panel in cyan), indicating that the $\text{LPML}_{\mathbf{WR|yt}}$ effectively selects the true model in this case. In conclusion, the $\text{LPML}_{\mathbf{WR|yt}}$ can serve as a good criterion for selecting a better model under these two different settings.

Similarly to Simulation Scenarios I and II, in Table 5 for Simulation Scenario III and Table 6 for Simulation Scenario IV, the biases and SDs of the posterior estimates are relatively small, SDs are similar to RMSEs, and most of the CPs are around the 95% nominal level. Due to the same reason (large missing percentages for the corresponding items) in Simulations I and II, the CP for b_{Y7} in Simulation Scenario III and CPs for b_{Y4} and b_{W11} in Simulation Scenario IV are lower than the nominal level. These CPs are improved when sample size increases, as shown in Table 10.

5.3. Simulation Scenarios V and VI

The purpose of simulations in this section is to investigate the impact of misspecification of the missing data model when the sizes of ηs are large ($\eta_T = \eta_W = \eta_R = -0.9$) in Simulation Scenario V and the sizes of ηs are small ($\eta_T = \eta_W = \eta_R = -0.1$) in Simulation Scenario VI. All else are the same as in Simulation Scenarios I and II. Under both settings, the missing data percentages are still controlled at 20%, with 15% due to omission behavior and 5% due to not-reached behavior.

Similarly, we fit both nonignorable and ignorable models to the simulated nonignorable datasets in Simulation Scenarios V and VI. The boxplots of the differences values $\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) under these two models are given in Figure 4. For Simulation Scenario V (Figure 4 left panel in red) where the true missing data mechanism is nonignorable and the sizes of ηs are large, the median of the values of $\text{LPML}_{\mathbf{WR|yt}}$ under the nonignorable model (−9470.20) is greater than that under the ignorable model (−9606.93). Additionally, the box (Median=133.33, IQR=(121.68, 145.16)) for the $\Delta\text{LPML}_{\mathbf{WR|yt}}$ is above 0, indicating that the nonignorable model outperforms the ignorable model when the true missing data mechanism is nonignorable and the size of η is large. For Simulation Scenario VI where the true missing data mechanism is nonignorable but the extent of nonignorability is small (i.e., the sizes of ηs are small), the median of the values of $\text{LPML}_{\mathbf{WR|yt}}$ under the nonignorable model (−9759.10) is slightly greater than that under the ignorable model (−9760.90). The difference $\Delta\text{LPML}_{\mathbf{WR|yt}}$ (Median=1.00, IQR= (−0.44, 1.57)) is not as significant as in Scenario V when the sizes of ηs are large, which is expected.

Similarly to Simulation Scenarios I and II, in Table 7 for Simulation Scenario V and Table 8 for Simulation Scenario VI, the biases and SDs of the posterior estimates are relatively small, SDs are similar

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 2. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the ignorable missing data mechanism in Simulation Scenario II (n = 2,000, J = 12, $\eta_W = \eta_R = 0$, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is ignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	0.019	0.070	0.072	0.947	b_{Y1}	−0.017	0.223	0.210	0.953
a_{Y2}	0.004	0.151	0.154	0.948	b_{Y2}	0.015	0.088	0.090	0.948
a_{Y3}	0.004	0.110	0.111	0.940	b_{Y3}	0.013	0.097	0.101	0.934
a_{Y4}	0.034	0.082	0.083	0.954	b_{Y4}	0.241	0.326	0.369	0.891
a_{Y5}	−0.003	0.086	0.090	0.940	b_{Y5}	0.014	0.080	0.081	0.941
a_{Y6}	0.015	0.074	0.071	0.954	b_{Y6}	−0.001	0.252	0.219	0.963
a_{Y7}	−0.010	0.118	0.122	0.949	b_{Y7}	0.011	0.065	0.067	0.939
a_{Y8}	0.002	0.098	0.100	0.935	b_{Y8}	0.015	0.095	0.102	0.932
a_{Y9}	−0.007	0.069	0.068	0.951	b_{Y9}	0.036	0.165	0.164	0.942
a_{Y10}	−0.013	0.079	0.081	0.933	b_{Y10}	0.022	0.093	0.093	0.941
a_{Y11}	0.078	0.206	0.223	0.942	b_{Y11}	−0.020	0.178	0.172	0.944
a_{Y12}	−0.004	0.094	0.091	0.955	b_{Y12}	0.021	0.098	0.099	0.952
a_{T1}	0.007	0.093	0.088	0.958	b_{T1}	0.000	0.017	0.018	0.940
a_{T2}	−0.002	0.092	0.088	0.955	b_{T2}	−0.001	0.018	0.018	0.945
a_{T3}	0.002	0.079	0.076	0.965	b_{T3}	0.000	0.018	0.018	0.937
a_{T4}	0.005	0.083	0.085	0.945	b_{T4}	0.000	0.029	0.030	0.937
a_{T5}	0.004	0.136	0.130	0.959	b_{T5}	0.000	0.016	0.016	0.951
a_{T6}	−0.001	0.103	0.104	0.944	b_{T6}	−0.002	0.020	0.020	0.950
a_{T7}	0.002	0.056	0.056	0.941	b_{T7}	0.000	0.022	0.022	0.943
a_{T8}	0.003	0.070	0.070	0.950	b_{T8}	−0.002	0.020	0.020	0.941
a_{T9}	−0.002	0.102	0.099	0.957	b_{T9}	−0.001	0.017	0.017	0.951
a_{T10}	0.002	0.093	0.095	0.934	b_{T10}	0.000	0.018	0.017	0.952
a_{T11}	0.003	0.083	0.082	0.946	b_{T11}	0.001	0.025	0.025	0.952
a_{T12}	0.004	0.100	0.101	0.939	b_{T12}	−0.001	0.018	0.018	0.950
b_{W1}	0.012	0.442	0.414	0.971	b_{R1}	0.003	0.092	0.096	0.943
b_{W2}	0.074	0.334	0.337	0.953	b_{R2}	0.006	0.070	0.071	0.945
b_{W3}	0.079	0.347	0.351	0.950	b_{R3}	−0.011	0.148	0.148	0.944
b_{W4}	0.045	0.127	0.137	0.941	b_{R4}	0.003	0.056	0.056	0.945
b_{W5}	0.077	0.221	0.243	0.935	b_{R5}	0.002	0.079	0.082	0.939
b_{W6}	0.088	0.248	0.264	0.945	b_{R6}	0.004	0.061	0.060	0.947
b_{W7}	0.036	0.432	0.411	0.965	b_{R7}	0.004	0.080	0.081	0.945
b_{W8}	0.101	0.337	0.352	0.957	b_{R8}	0.006	0.089	0.089	0.950
b_{W9}	0.094	0.255	0.287	0.944	b_{R9}	0.006	0.097	0.101	0.933
b_{W10}	0.059	0.386	0.385	0.955	b_{R10}	−0.005	0.110	0.111	0.945
b_{W11}	−0.781	0.723	0.876	0.841	b_{R11}	0.005	0.061	0.063	0.944
b_{W12}	0.003	0.075	0.073	0.949	σ_T	0.000	0.008	0.008	0.947

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 3. Bias, SD, RMSE, and CP of parameters in the item response model and the response time model under MICE in Simulation Scenario I (n = 2,000, J = 12, median sizes of ηs, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.022	0.094	0.100	0.932	b_{Y1}	−0.037	0.117	0.126	0.911
a_{Y2}	−0.003	0.077	0.076	0.952	b_{Y2}	−0.025	0.137	0.133	0.955
a_{Y3}	−0.005	0.072	0.070	0.947	b_{Y3}	−0.009	0.246	0.229	0.954
a_{Y4}	0.010	0.110	0.113	0.945	b_{Y4}	−0.012	0.159	0.160	0.948
a_{Y5}	−0.007	0.092	0.092	0.948	b_{Y5}	−0.016	0.163	0.159	0.944
a_{Y6}	0.012	0.070	0.065	0.964	b_{Y6}	−0.012	0.231	0.204	0.967
a_{Y7}	0.032	0.080	0.084	0.938	b_{Y7}	−0.016	0.383	0.335	0.952
a_{Y8}	−0.014	0.083	0.083	0.950	b_{Y8}	−0.038	0.124	0.125	0.938
a_{Y9}	−0.021	0.120	0.118	0.946	b_{Y9}	−0.043	0.146	0.146	0.933
a_{Y10}	−0.019	0.107	0.111	0.939	b_{Y10}	−0.031	0.084	0.088	0.933
a_{Y11}	−0.014	0.094	0.093	0.949	b_{Y11}	−0.006	0.126	0.121	0.955
a_{Y12}	−0.010	0.089	0.090	0.948	b_{Y12}	−0.040	0.109	0.109	0.952
a_{T1}	−0.048	0.102	0.127	0.899	b_{T1}	0.009	0.019	0.021	0.924
a_{T2}	−0.021	0.087	0.091	0.931	b_{T2}	0.003	0.017	0.018	0.938
a_{T3}	−0.035	0.086	0.098	0.913	b_{T3}	0.007	0.019	0.020	0.938
a_{T4}	−0.009	0.057	0.057	0.945	b_{T4}	0.002	0.021	0.021	0.949
a_{T5}	−0.056	0.135	0.157	0.910	b_{T5}	0.004	0.016	0.017	0.937
a_{T6}	−0.016	0.087	0.088	0.949	b_{T6}	0.000	0.017	0.017	0.952
a_{T7}	−0.068	0.073	0.105	0.831	b_{T7}	0.019	0.030	0.037	0.894
a_{T8}	−0.018	0.071	0.075	0.929	b_{T8}	0.004	0.020	0.021	0.947
a_{T9}	−0.064	0.113	0.142	0.889	b_{T9}	0.009	0.019	0.021	0.926
a_{T10}	−0.026	0.096	0.105	0.921	b_{T10}	0.004	0.018	0.019	0.952
a_{T11}	−0.014	0.065	0.066	0.943	b_{T11}	0.003	0.021	0.021	0.955
a_{T12}	−0.016	0.099	0.102	0.940	b_{T12}	0.001	0.018	0.019	0.952
σ_T	−0.010	0.008	0.016	0.783					
η_T	0.009	0.038	0.039	0.937					

to RMSEs, and most of the CPs are around the 95% nominal level. Due to the same reason (large missing percentages for the corresponding items) in Simulations I and II, the CP for b_{Y7} in Simulation Scenario V and CPs for b_{Y4} and b_{W11} in Simulation Scenario VI are lower than the nominal level. These CPs are improved when sample size increases, as shown in Table 10.

5.4. Simulation Scenarios VII and VIII

The purpose of simulations in this section is to investigate the impact of misspecification of the missing data model when the sample size is small (n = 500) in Simulation Scenario VII and large (n = 4,000) in Simulation Scenario VIII, with all other conditions identical to Scenario I.

Similarly, we fit both nonignorable and ignorable models to the simulated nonignorable datasets in Simulation Scenario VII and the simulated ignorable datasets in Simulation Scenario VIII. The boxplots

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 4. Bias, SD, RMSE, and CP of parameters in the item response model and the response time model under the response time process model in Simulation Scenario I (n = 2,000, J = 12, median sizes of ηs, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.022	0.095	0.100	0.937	b_{Y1}	−0.056	0.118	0.132	0.891
a_{Y2}	−0.013	0.079	0.079	0.940	b_{Y2}	−0.061	0.149	0.154	0.941
a_{Y3}	−0.004	0.072	0.068	0.952	b_{Y3}	−0.060	0.252	0.240	0.927
a_{Y4}	0.005	0.112	0.115	0.941	b_{Y4}	−0.003	0.163	0.164	0.944
a_{Y5}	−0.007	0.094	0.093	0.945	b_{Y5}	−0.018	0.166	0.162	0.934
a_{Y6}	0.009	0.071	0.065	0.969	b_{Y6}	0.002	0.240	0.209	0.964
a_{Y7}	0.032	0.082	0.075	0.969	b_{Y7}	−0.066	0.387	0.295	0.978
a_{Y8}	−0.018	0.083	0.084	0.944	b_{Y8}	−0.061	0.128	0.136	0.913
a_{Y9}	−0.018	0.123	0.119	0.947	b_{Y9}	−0.048	0.149	0.149	0.931
a_{Y10}	−0.020	0.108	0.112	0.936	b_{Y10}	−0.039	0.085	0.092	0.918
a_{Y11}	−0.015	0.095	0.095	0.944	b_{Y11}	−0.008	0.127	0.123	0.948
a_{Y12}	−0.016	0.090	0.093	0.938	b_{Y12}	−0.057	0.113	0.120	0.926
a_{T1}	−0.415	0.086	0.424	0.011	b_{T1}	0.205	0.019	0.206	0.000
a_{T2}	−0.125	0.082	0.149	0.659	b_{T2}	0.047	0.017	0.050	0.212
a_{T3}	−0.268	0.076	0.279	0.085	b_{T3}	0.147	0.019	0.149	0.000
a_{T4}	−0.074	0.054	0.091	0.722	b_{T4}	0.057	0.021	0.060	0.206
a_{T5}	−0.340	0.122	0.360	0.236	b_{T5}	0.092	0.016	0.093	0.000
a_{T6}	−0.070	0.085	0.108	0.865	b_{T6}	0.013	0.017	0.022	0.878
a_{T7}	−0.527	0.050	0.529	0.000	b_{T7}	0.667	0.032	0.668	0.000
a_{T8}	−0.135	0.066	0.150	0.477	b_{T8}	0.087	0.020	0.089	0.009
a_{T9}	−0.436	0.096	0.446	0.009	b_{T9}	0.191	0.019	0.192	0.000
a_{T10}	−0.179	0.089	0.200	0.476	b_{T10}	0.077	0.018	0.079	0.013
a_{T11}	−0.088	0.062	0.107	0.706	b_{T11}	0.056	0.021	0.059	0.236
a_{T12}	−0.048	0.098	0.110	0.906	b_{T12}	−0.002	0.019	0.019	0.942
σ_T	0.073	0.038	0.083	0.489					
η_T	−0.008	0.008	0.011	0.850					

of the differences values $\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) under these two models are given in Figure 5. All the three boxplots are above 0, indicating that the nonignorable model outperforms the ignorable model when the true missing data mechanism is nonignorable. As sample size increases, the power of $\text{LPML}_{\mathbf{WR|yt}}$ becomes greater, which is visualized by the boxplots diverging further from the reference line $y = 0$. Specifically, the median (IQR) values are 63.86 (55.90, 72.34) for Scenario I, 14.05 (10.33, 18.23) for Scenario VII, and 129.94 (119.49, 140.93) for Scenario VIII.

Similarly to Simulation Scenario I, in Table 9 for Simulation Scenario VII, the biases and SDs of the posterior estimates are relatively small, SDs are similar to RMSEs, and most of the CPs are around the 95% nominal level. Results for CPs are further improved in Table 10 when the sample size is large, with all CPs being around the nominal level. Additionally, SDs and RMSEs in Simulation Scenario VIII (n = 4,000) are the smallest, followed by those in Simulation Scenario I (n = 2,000).

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

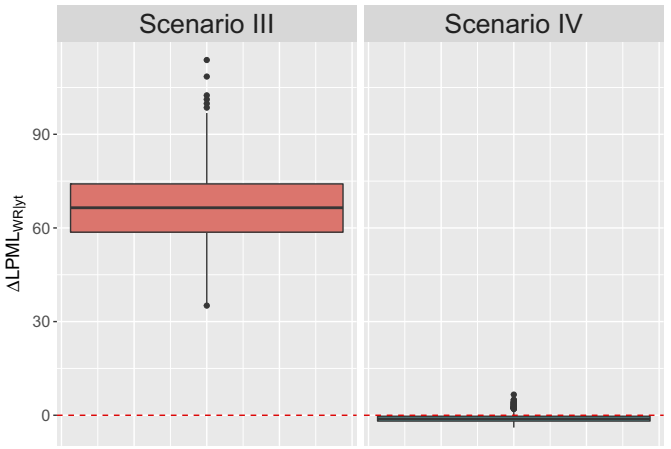


Figure 3. Boxplots of $\Delta\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) when the true missing data mechanisms are nonignorable (Simulation Scenario III in red) and ignorable (Simulation Scenario IV in cyan).

5.5. Simulation Scenario IX

The purpose of simulations in this section is to investigate the impact of misspecification of the missing data model when the missing data percentages of the simulated datasets are around 20% (5% due to omission behavior and 15% due to not-reached behavior). The proportions of missing percentage are different from those in Simulation Scenario I (15% is due to omission and 5% is due to not-reached behavior). Other conditions are identical to Scenario I.

We fit both nonignorable and ignorable models to the simulated nonignorable datasets in Simulation Scenario IX. The boxplot of difference values $\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) under these two models is given in Figure 6. All the two boxplots are above 0, indicating that the nonignorable model outperforms the ignorable model when the true missing data mechanism is nonignorable. The power of $\text{LPML}_{\mathbf{WR|yt}}$ in Simulation Scenario I is greater than that in Simulation Scenario IX, which is visualized by the boxplots diverging further from the reference line $y = 0$. Specifically, the median (IQR) values are 63.86 (55.90, 72.34) for Scenario I and 39.20 (33.58, 45.33) for Scenario IX.

Similarly to Simulation Scenario I, in Table 11 for Simulation Scenario IX, the biases and SDs of the posterior estimates are relatively small, SDs are similar to RMSEs, and most of the CPs are around the 95% nominal level. SDs and RMSEs of W model in Simulation Scenario IX are smaller than those in Simulation Scenario I. SDs and RMSEs of R model in Simulation Scenario IX are greater than those in Simulation Scenario I. Both results are expected given that we have more missingness due to not-reached and less missingness due to omission in Scenario IX.

Posterior summaries remain good for the parameters of R model even when the percentage of missingness due to omission is small (5%) in Simulation Scenario IX and for the parameters of W model even when the percentage of missingness due to not-reached is small (5%) in Simulation Scenario I. This indicates that our methods perform robustly when the missing data percentages for both omission and not-reached are small.

6. Analysis of the PISA 2018 Test Data

6.1. Data overview

The PISA 2018 math exam in Brazil includes six test booklets. The percentages of missing items for each booklet are 18.6%, 10.9%, 13.0%, 14.1%, 13.6%, and 16.8%, respectively. In our real data analysis, we focus on the first test, which comprises 1,742 records ($n = 1{,}742$) and 12 items ($J = 12$). The total time limit is 60 minutes. The percentage of missing items ($\mathbf{R}$) is 18.6% (12.3% due to omission and 6.3% due to

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 5. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the nonignorable missing data mechanism in Simulation Scenario III ($n = 2,000$, $J = 12$, median sizes of ηs, 30% missing data percentage (20% due to omission + 10% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.021	0.101	0.102	0.940	b_{Y1}	0.032	0.123	0.123	0.949
a_{Y2}	0.003	0.077	0.075	0.953	b_{Y2}	0.013	0.136	0.130	0.948
a_{Y3}	0.000	0.075	0.069	0.969	b_{Y3}	0.049	0.250	0.220	0.956
a_{Y4}	0.026	0.111	0.113	0.941	b_{Y4}	−0.008	0.153	0.146	0.958
a_{Y5}	0.007	0.095	0.092	0.949	b_{Y5}	0.018	0.160	0.152	0.958
a_{Y6}	0.015	0.069	0.065	0.956	b_{Y6}	−0.012	0.223	0.200	0.962
a_{Y7}	0.064	0.094	0.101	0.957	b_{Y7}	0.347	0.346	0.439	0.854
a_{Y8}	−0.010	0.086	0.085	0.945	b_{Y8}	0.016	0.131	0.124	0.954
a_{Y9}	0.012	0.130	0.125	0.951	b_{Y9}	0.012	0.147	0.143	0.946
a_{Y10}	−0.019	0.110	0.110	0.953	b_{Y10}	0.016	0.087	0.087	0.949
a_{Y11}	0.002	0.096	0.094	0.953	b_{Y11}	0.017	0.124	0.121	0.944
a_{Y12}	−0.007	0.088	0.090	0.941	b_{Y12}	0.002	0.113	0.109	0.937
a_{T1}	0.011	0.113	0.116	0.944	b_{T1}	−0.001	0.020	0.020	0.958
a_{T2}	0.003	0.090	0.093	0.941	b_{T2}	0.000	0.018	0.018	0.942
a_{T3}	0.001	0.093	0.093	0.948	b_{T3}	0.000	0.021	0.020	0.948
a_{T4}	0.003	0.059	0.059	0.943	b_{T4}	0.000	0.022	0.021	0.944
a_{T5}	0.001	0.146	0.143	0.949	b_{T5}	−0.001	0.017	0.017	0.952
a_{T6}	0.008	0.091	0.087	0.960	b_{T6}	−0.002	0.018	0.018	0.946
a_{T7}	0.004	0.090	0.091	0.950	b_{T7}	0.000	0.035	0.035	0.952
a_{T8}	−0.002	0.076	0.076	0.944	b_{T8}	−0.001	0.022	0.021	0.946
a_{T9}	0.004	0.127	0.122	0.958	b_{T9}	0.000	0.021	0.021	0.944
a_{T10}	−0.002	0.103	0.103	0.944	b_{T10}	0.000	0.020	0.020	0.943
a_{T11}	0.001	0.069	0.067	0.953	b_{T11}	0.001	0.022	0.021	0.951
a_{T12}	−0.001	0.103	0.100	0.950	b_{T12}	−0.001	0.019	0.019	0.953
b_{W1}	−0.039	0.420	0.401	0.954	b_{R1}	0.003	0.056	0.055	0.946
b_{W2}	−0.078	0.501	0.460	0.955	b_{R2}	0.001	0.080	0.084	0.939
b_{W3}	0.031	0.115	0.119	0.943	b_{R3}	0.003	0.061	0.063	0.939
b_{W4}	0.040	0.380	0.384	0.955	b_{R4}	−0.003	0.080	0.079	0.951
b_{W5}	0.059	0.282	0.297	0.949	b_{R5}	0.003	0.065	0.067	0.942
b_{W6}	0.038	0.290	0.293	0.948	b_{R6}	−0.011	0.126	0.128	0.941
b_{W7}	0.024	0.109	0.114	0.937	b_{R7}	0.006	0.060	0.059	0.949
b_{W8}	0.053	0.190	0.203	0.937	b_{R8}	0.004	0.073	0.075	0.932
b_{W9}	0.051	0.212	0.220	0.952	b_{R9}	0.003	0.060	0.060	0.950
b_{W10}	0.058	0.383	0.388	0.960	b_{R10}	0.002	0.073	0.074	0.935
b_{W11}	0.059	0.290	0.300	0.940	b_{R11}	0.004	0.086	0.087	0.942

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 5. Continued

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
b_{W12}	0.004	0.079	0.081	0.946	σ_T	0.000	0.009	0.009	0.944
η_T	0.003	0.038	0.039	0.931					
η_W	0.021	0.104	0.115	0.910					
α_W	0.066	0.376	0.241	0.971					
η_R	0.000	0.045	0.046	0.930					
α_R	0.006	0.091	0.090	0.950					

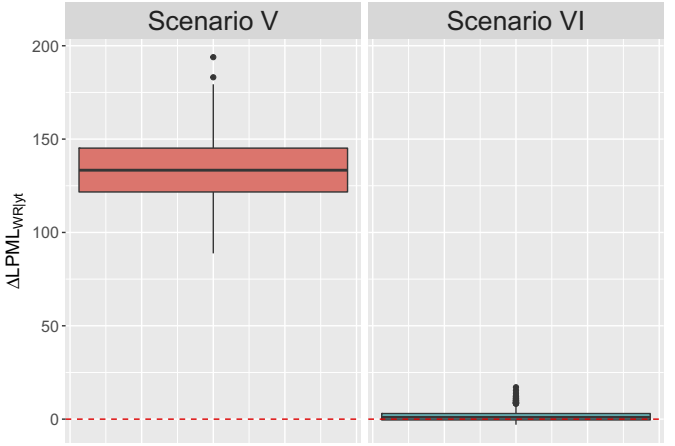


Figure 4. Boxplots of $\Delta\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ (Nonignorable–Ignorable) when the true missing data mechanisms are nonignorable with large sizes of ηs (Simulation Scenario V in red) and nonignorable with small sizes of ηs (Simulation Scenario VI in cyan).

not-reached) and the percentage of not-reached behavior (**W**) is 18.7%. Table 12 shows the not-reached and omitted behaviors, where the (i,j)th element represents the total number of missing answers for item j among subjects who fail to reach item i. The lower triangular matrix in red and bold font then shows the number of omitted items for item j among subjects who fail to reach item i, for $j \leq i-2$. Each row total for $W \leq 12$ returns the total number of subjects that fail to reach the corresponding item, which is the same as the corresponding diagonal element. Each column total returns the total number of missing answers for that item, with the total number of omitted items in parentheses and in red and bold font. Table 13 presents a summary of the response accuracy and response times for each item.

6.2. Posterior estimation

We fit the data using the proposed item response and response time models proposed in Section 2, together with both nonignorable and ignorable (i.e., set $\eta_W = \eta_R = 0$) missing data mechanisms proposed in Section 3. We evaluate the convergence of models using traceplots and autocorrelation plots, as shown in the Supplementary Material. All plots indicate good mixing of the chains and convergence for all parameters. We define a posterior estimate to be "statistically significant at a significance level of 0.05" if the corresponding 95% HPD interval does not include 0.

Table 14 presents the posterior summaries of the difficulty parameters in the response model, that is, $a_{Yj}b_{Yj}$, where a larger value indicates that the j^{th} item is more difficult and its answer is more likely wrong. For example, $a_{Y1}b_{Y1} = -0.759$ (95% HPD = $(-0.883, -0.642)$) is the smallest significant value reported in Table 14, which is consistent with the fact that Item 1 has the highest percentage of correct answers as shown in Table 13. $a_{Y9}b_{Y9} = 5.062$ (95% HPD = $(4.400, 45.958)$) and $a_{Y11}b_{Y11} = 5.366$ (95% HPD =

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 6. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the ignorable missing data mechanism in Simulation Scenario IV (n = 2,000, J = 12, $\eta_W = \eta_R = 0$, 30% missing data percentage (20% due to omission + 10% due to not-reached), and the true missing data mechanism is ignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	0.022	0.073	0.069	0.960	b_{Y1}	−0.024	0.229	0.205	0.963
a_{Y2}	0.007	0.164	0.163	0.936	b_{Y2}	0.015	0.094	0.094	0.944
a_{Y3}	0.005	0.115	0.121	0.935	b_{Y3}	0.015	0.101	0.104	0.940
a_{Y4}	0.041	0.093	0.092	0.960	b_{Y4}	0.332	0.358	0.453	0.847
a_{Y5}	−0.005	0.091	0.088	0.957	b_{Y5}	0.013	0.084	0.083	0.957
a_{Y6}	0.023	0.081	0.079	0.961	b_{Y6}	−0.010	0.267	0.233	0.977
a_{Y7}	−0.008	0.127	0.129	0.933	b_{Y7}	0.010	0.068	0.067	0.950
a_{Y8}	0.001	0.105	0.107	0.945	b_{Y8}	0.017	0.101	0.105	0.938
a_{Y9}	−0.003	0.073	0.070	0.948	b_{Y9}	0.046	0.173	0.171	0.930
a_{Y10}	−0.009	0.083	0.089	0.923	b_{Y10}	0.015	0.097	0.099	0.947
a_{Y11}	0.097	0.234	0.243	0.955	b_{Y11}	−0.032	0.195	0.188	0.949
a_{Y12}	−0.008	0.098	0.095	0.951	b_{Y12}	0.026	0.102	0.103	0.948
a_{T1}	0.005	0.095	0.091	0.958	b_{T1}	0.000	0.017	0.017	0.951
a_{T2}	−0.002	0.096	0.097	0.944	b_{T2}	−0.001	0.019	0.019	0.945
a_{T3}	0.002	0.080	0.078	0.945	b_{T3}	0.001	0.018	0.018	0.930
a_{T4}	0.003	0.095	0.097	0.947	b_{T4}	0.001	0.033	0.032	0.956
a_{T5}	0.006	0.144	0.150	0.933	b_{T5}	0.000	0.017	0.017	0.934
a_{T6}	−0.002	0.112	0.113	0.948	b_{T6}	−0.001	0.021	0.021	0.955
a_{T7}	0.004	0.059	0.060	0.937	b_{T7}	0.001	0.023	0.023	0.946
a_{T8}	0.000	0.074	0.073	0.948	b_{T8}	−0.001	0.021	0.021	0.947
a_{T9}	0.007	0.108	0.111	0.939	b_{T9}	0.000	0.018	0.018	0.946
a_{T10}	0.001	0.097	0.099	0.939	b_{T10}	0.000	0.018	0.019	0.939
a_{T11}	0.003	0.093	0.092	0.950	b_{T11}	0.001	0.028	0.029	0.945
a_{T12}	−0.001	0.102	0.102	0.947	b_{T12}	0.000	0.019	0.019	0.935
b_{W1}	0.013	0.353	0.356	0.950	b_{R1}	0.009	0.079	0.080	0.937
b_{W2}	0.051	0.264	0.273	0.951	b_{R2}	0.003	0.063	0.065	0.942
b_{W3}	0.062	0.275	0.280	0.948	b_{R3}	0.001	0.123	0.117	0.950
b_{W4}	0.027	0.104	0.107	0.939	b_{R4}	0.005	0.059	0.061	0.939
b_{W5}	0.056	0.180	0.191	0.948	b_{R5}	0.002	0.071	0.072	0.947
b_{W6}	0.057	0.203	0.216	0.948	b_{R6}	0.002	0.059	0.060	0.933
b_{W7}	0.015	0.355	0.347	0.954	b_{R7}	0.003	0.071	0.074	0.924
b_{W8}	0.058	0.275	0.284	0.955	b_{R8}	0.003	0.079	0.076	0.950
b_{W9}	0.056	0.209	0.218	0.943	b_{R9}	0.001	0.084	0.084	0.952
b_{W10}	0.052	0.324	0.332	0.948	b_{R10}	0.006	0.095	0.096	0.936
b_{W11}	−0.616	0.666	0.744	0.870	b_{R11}	0.003	0.063	0.064	0.936
b_{W12}	0.002	0.078	0.080	0.937	σ_T	0.000	0.009	0.009	0.947

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 7. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the nonignorable missing data mechanism in Simulation Scenario V (n = 2,000, J = 12, large sizes of ηs, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.004	0.072	0.073	0.946	b_{Y1}	0.000	0.103	0.098	0.958
a_{Y2}	0.003	0.059	0.056	0.957	b_{Y2}	0.017	0.121	0.113	0.955
a_{Y3}	−0.001	0.059	0.057	0.954	b_{Y3}	0.043	0.221	0.205	0.952
a_{Y4}	0.019	0.081	0.082	0.947	b_{Y4}	−0.012	0.120	0.121	0.944
a_{Y5}	0.007	0.071	0.069	0.950	b_{Y5}	0.004	0.128	0.127	0.947
a_{Y6}	0.015	0.055	0.055	0.953	b_{Y6}	−0.023	0.188	0.181	0.930
a_{Y7}	0.040	0.070	0.072	0.957	b_{Y7}	0.264	0.320	0.369	0.894
a_{Y8}	−0.002	0.064	0.066	0.941	b_{Y8}	0.004	0.116	0.115	0.945
a_{Y9}	0.011	0.091	0.091	0.949	b_{Y9}	−0.002	0.111	0.111	0.940
a_{Y10}	0.003	0.078	0.081	0.938	b_{Y10}	0.002	0.073	0.074	0.944
a_{Y11}	0.004	0.071	0.070	0.962	b_{Y11}	0.006	0.101	0.103	0.944
a_{Y12}	0.000	0.068	0.073	0.937	b_{Y12}	0.003	0.105	0.105	0.944
a_{T1}	0.000	0.103	0.103	0.945	b_{T1}	0.000	0.019	0.019	0.951
a_{T2}	−0.003	0.086	0.084	0.952	b_{T2}	0.000	0.018	0.018	0.929
a_{T3}	0.001	0.087	0.088	0.944	b_{T3}	0.001	0.019	0.019	0.955
a_{T4}	0.001	0.057	0.056	0.947	b_{T4}	0.000	0.021	0.021	0.939
a_{T5}	0.000	0.136	0.138	0.944	b_{T5}	0.001	0.016	0.016	0.959
a_{T6}	−0.001	0.087	0.090	0.942	b_{T6}	0.000	0.017	0.018	0.930
a_{T7}	0.002	0.077	0.078	0.941	b_{T7}	0.001	0.030	0.030	0.946
a_{T8}	−0.001	0.071	0.071	0.943	b_{T8}	−0.001	0.020	0.021	0.941
a_{T9}	−0.002	0.114	0.111	0.952	b_{T9}	0.000	0.019	0.019	0.953
a_{T10}	0.001	0.095	0.096	0.951	b_{T10}	0.000	0.019	0.018	0.950
a_{T11}	0.000	0.065	0.066	0.945	b_{T11}	0.001	0.021	0.021	0.946
a_{T12}	0.001	0.099	0.100	0.946	b_{T12}	0.001	0.019	0.019	0.951
b_{W1}	−0.023	0.535	0.468	0.962	b_{R1}	0.007	0.059	0.062	0.942
b_{W2}	−0.112	0.619	0.520	0.957	b_{R2}	0.003	0.094	0.098	0.935
b_{W3}	0.051	0.142	0.153	0.940	b_{R3}	0.008	0.066	0.068	0.938
b_{W4}	0.070	0.477	0.467	0.958	b_{R4}	0.005	0.095	0.095	0.943
b_{W5}	0.098	0.355	0.368	0.953	b_{R5}	0.006	0.072	0.074	0.939
b_{W6}	0.081	0.364	0.394	0.940	b_{R6}	−0.008	0.156	0.153	0.955
b_{W7}	0.032	0.130	0.133	0.943	b_{R7}	0.006	0.057	0.055	0.953
b_{W8}	0.077	0.230	0.251	0.933	b_{R8}	0.013	0.083	0.085	0.937
b_{W9}	0.095	0.259	0.281	0.943	b_{R9}	0.005	0.063	0.062	0.956
b_{W10}	0.047	0.454	0.450	0.956	b_{R10}	0.006	0.083	0.085	0.946
b_{W11}	0.097	0.352	0.381	0.945	b_{R11}	0.003	0.103	0.104	0.948

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 7. Continued

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
b_{W12}	0.002	0.076	0.077	0.938	σ_T	0.001	0.008	0.008	0.956
η_T	0.006	0.021	0.022	0.932					
η_W	0.027	0.069	0.062	0.976					
α_W	0.019	0.092	0.093	0.954					
η_R	0.015	0.056	0.052	0.956					
α_R	0.017	0.067	0.068	0.941					

Table 8. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the nonignorable missing data mechanism in Simulation Scenario VI ($n = 2{,}000$, $J = 12$, small sizes of ηs, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.018	0.099	0.100	0.939	b_{Y1}	0.030	0.126	0.126	0.947
a_{Y2}	0.001	0.080	0.080	0.953	b_{Y2}	0.005	0.134	0.124	0.956
a_{Y3}	−0.001	0.075	0.067	0.960	b_{Y3}	0.071	0.246	0.214	0.962
a_{Y4}	0.021	0.117	0.114	0.954	b_{Y4}	0.004	0.165	0.154	0.953
a_{Y5}	0.015	0.097	0.097	0.956	b_{Y5}	0.011	0.168	0.164	0.943
a_{Y6}	0.019	0.072	0.067	0.964	b_{Y6}	−0.008	0.231	0.208	0.958
a_{Y7}	0.044	0.088	0.087	0.965	b_{Y7}	0.246	0.331	0.364	0.911
a_{Y8}	−0.017	0.086	0.087	0.940	b_{Y8}	0.019	0.127	0.125	0.948
a_{Y9}	0.014	0.129	0.132	0.947	b_{Y9}	0.014	0.157	0.153	0.946
a_{Y10}	−0.005	0.114	0.116	0.945	b_{Y10}	0.014	0.089	0.088	0.954
a_{Y11}	0.000	0.099	0.098	0.947	b_{Y11}	0.019	0.130	0.132	0.945
a_{Y12}	−0.009	0.093	0.094	0.938	b_{Y12}	0.004	0.109	0.104	0.950
a_{T1}	−0.005	0.104	0.106	0.936	b_{T1}	0.000	0.019	0.019	0.945
a_{T2}	−0.001	0.087	0.089	0.936	b_{T2}	0.000	0.018	0.018	0.942
a_{T3}	−0.005	0.087	0.086	0.946	b_{T3}	0.000	0.020	0.019	0.951
a_{T4}	0.001	0.057	0.058	0.941	b_{T4}	0.000	0.021	0.021	0.955
a_{T5}	0.001	0.138	0.137	0.941	b_{T5}	−0.001	0.016	0.017	0.948
a_{T6}	0.001	0.088	0.086	0.947	b_{T6}	−0.001	0.017	0.017	0.943
a_{T7}	0.002	0.079	0.078	0.961	b_{T7}	0.001	0.030	0.030	0.957
a_{T8}	0.000	0.072	0.067	0.959	b_{T8}	−0.002	0.020	0.020	0.950
a_{T9}	0.003	0.116	0.115	0.940	b_{T9}	−0.001	0.020	0.020	0.942
a_{T10}	0.002	0.097	0.097	0.946	b_{T10}	−0.001	0.019	0.019	0.933
a_{T11}	0.002	0.066	0.064	0.951	b_{T11}	0.000	0.021	0.021	0.937
a_{T12}	0.005	0.100	0.100	0.947	b_{T12}	−0.001	0.019	0.018	0.942
b_{W1}	0.005	0.545	0.503	0.967	b_{R1}	0.007	0.059	0.062	0.943

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 8. Continued

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
b_{W2}	−0.123	0.619	0.543	0.960	b_{R2}	0.003	0.093	0.094	0.937
b_{W3}	0.038	0.142	0.149	0.941	b_{R3}	0.007	0.065	0.068	0.940
b_{W4}	0.075	0.478	0.460	0.968	b_{R4}	0.003	0.093	0.093	0.944
b_{W5}	0.100	0.356	0.392	0.945	b_{R5}	0.004	0.071	0.069	0.955
b_{W6}	0.093	0.366	0.386	0.949	b_{R6}	−0.013	0.151	0.148	0.950
b_{W7}	0.037	0.130	0.136	0.935	b_{R7}	0.006	0.057	0.057	0.938
b_{W8}	0.076	0.230	0.251	0.937	b_{R8}	0.003	0.080	0.081	0.953
b_{W9}	0.085	0.257	0.277	0.940	b_{R9}	0.004	0.062	0.064	0.941
b_{W10}	0.057	0.455	0.452	0.968	b_{R10}	0.007	0.081	0.084	0.943
b_{W11}	0.109	0.354	0.382	0.954	b_{R11}	0.004	0.097	0.094	0.951
b_{W12}	0.002	0.076	0.076	0.940	σ_T	0.000	0.008	0.009	0.936
η_T	−0.002	0.039	0.038	0.942					
η_W	0.038	0.112	0.106	0.959					
α_W	0.676	0.785	0.733	0.979					
η_R	0.026	0.062	0.067	0.924					
α_R	0.427	0.670	0.576	0.974					

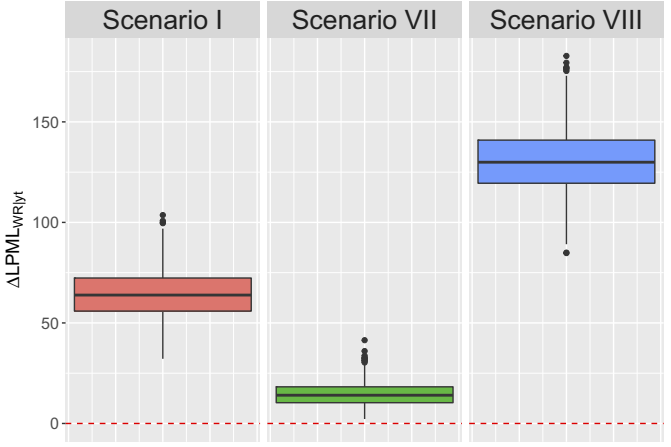


Figure 5. Boxplots of $\Delta\text{LPML}_{\mathbf{WR|yt}}$ (Nonignorable–Ignorable) when the true missing data mechanisms are nonignorable with $n = 2{,}000$ (Simulation Scenario I = in red), nonignorable with $n = 500$ (Simulation Scenario VII in green), and nonignorable with $n = 4{,}000$ (Simulation Scenario VIII in cyan).

(4.629,6.300)) are the first and second highest values, which aligns well with the fact that these two items have the smallest percentages of correctness in Table 13.

According to Table 15, $b_{T3} = 3.125$(95% HPD = (3.066,3.186)), $b_{T7} = 3.088$ (95% HPD = (3.036,3.147)), and $b_{T10} = 3.209$ (95% HPD = (3.179,3.242)) are the three significant parameters with the highest values among all items. This aligns well with the findings in Table 13, where students tend to spend more time on Items 3, 7, and 10. $b_{R3} = 0.872$ (95% HPD = (0.749,1.008)) and $b_{R5} = 0.584$ (95% HPD = (0.457,0.718)) are the two significant parameters with the smallest values among all items, which is consistent with the fact that the total numbers of omitted missingness of Items 3 and 5 are the

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 9. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the nonignorable missing data mechanism in Simulation Scenario VII (n = 500, J = 12, median sizes of ηs, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.037	0.176	0.171	0.951	b_{Y1}	0.066	0.262	0.234	0.955
a_{Y2}	0.010	0.141	0.133	0.961	b_{Y2}	0.108	0.269	0.261	0.929
a_{Y3}	0.043	0.122	0.112	0.978	b_{Y3}	0.119	0.387	0.317	0.967
a_{Y4}	0.095	0.210	0.219	0.958	b_{Y4}	−0.061	0.266	0.241	0.944
a_{Y5}	0.039	0.170	0.163	0.964	b_{Y5}	−0.004	0.283	0.236	0.964
a_{Y6}	0.056	0.122	0.122	0.957	b_{Y6}	−0.095	0.340	0.271	0.964
a_{Y7}	0.112	0.148	0.161	0.964	b_{Y7}	0.575	0.456	0.663	0.797
a_{Y8}	−0.015	0.151	0.154	0.942	b_{Y8}	0.081	0.281	0.247	0.948
a_{Y9}	0.054	0.233	0.224	0.960	b_{Y9}	−0.001	0.262	0.227	0.957
a_{Y10}	−0.034	0.205	0.206	0.941	b_{Y10}	0.060	0.189	0.181	0.958
a_{Y11}	0.006	0.175	0.165	0.957	b_{Y11}	0.040	0.244	0.214	0.965
a_{Y12}	−0.022	0.164	0.166	0.938	b_{Y12}	0.056	0.253	0.226	0.956
a_{T1}	−0.009	0.207	0.206	0.939	b_{T1}	0.003	0.039	0.038	0.955
a_{T2}	0.005	0.175	0.175	0.945	b_{T2}	0.001	0.035	0.035	0.948
a_{T3}	0.006	0.176	0.170	0.946	b_{T3}	0.002	0.039	0.037	0.957
a_{T4}	0.003	0.115	0.114	0.950	b_{T4}	0.001	0.042	0.042	0.948
a_{T5}	−0.001	0.275	0.272	0.943	b_{T5}	−0.001	0.033	0.032	0.952
a_{T6}	0.009	0.177	0.183	0.929	b_{T6}	−0.003	0.035	0.034	0.944
a_{T7}	0.016	0.159	0.157	0.951	b_{T7}	0.000	0.061	0.061	0.940
a_{T8}	0.010	0.144	0.144	0.955	b_{T8}	−0.003	0.041	0.040	0.955
a_{T9}	0.001	0.232	0.235	0.940	b_{T9}	−0.001	0.039	0.040	0.942
a_{T10}	0.004	0.194	0.197	0.935	b_{T10}	0.000	0.037	0.038	0.943
a_{T11}	0.006	0.132	0.136	0.945	b_{T11}	0.000	0.042	0.042	0.937
a_{T12}	0.003	0.200	0.198	0.946	b_{T12}	0.000	0.038	0.037	0.953
b_{W1}	−0.148	0.950	0.694	0.956	b_{R1}	0.009	0.118	0.120	0.932
b_{W2}	−0.379	1.013	0.761	0.936	b_{R2}	−0.007	0.185	0.179	0.964
b_{W3}	0.178	0.299	0.367	0.911	b_{R3}	0.013	0.131	0.128	0.952
b_{W4}	0.031	0.882	0.717	0.975	b_{R4}	0.006	0.187	0.186	0.950
b_{W5}	0.242	0.725	0.755	0.973	b_{R5}	0.014	0.142	0.153	0.936
b_{W6}	0.242	0.748	0.761	0.975	b_{R6}	−0.060	0.297	0.297	0.935
b_{W7}	0.167	0.276	0.345	0.901	b_{R7}	0.020	0.114	0.112	0.951
b_{W8}	0.283	0.486	0.623	0.914	b_{R8}	0.006	0.161	0.158	0.956
b_{W9}	0.299	0.539	0.641	0.930	b_{R9}	0.014	0.125	0.126	0.944
b_{W10}	0.128	0.878	0.740	0.979	b_{R10}	0.004	0.162	0.159	0.956
b_{W11}	0.287	0.723	0.771	0.968	b_{R11}	0.013	0.195	0.194	0.948

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 9. Continued

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
b_{W12}	0.003	0.152	0.153	0.948	σ_T	0.002	0.017	0.017	0.943
η_T	0.012	0.073	0.077	0.923					
η_W	0.064	0.209	0.197	0.963					
α_W	0.145	0.691	0.297	0.990					
η_R	0.011	0.089	0.093	0.935					
α_R	0.052	0.200	0.195	0.948					

Table 10. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the nonignorable missing data mechanism in Simulation Scenario VIII ($n = 4,000$, $J = 12$, median sizes of ηs, 20% missing data percentage (15% due to omission + 5% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.006	0.065	0.065	0.939	b_{Y1}	0.012	0.076	0.075	0.954
a_{Y2}	0.002	0.053	0.052	0.941	b_{Y2}	0.005	0.091	0.089	0.949
a_{Y3}	−0.005	0.053	0.051	0.946	b_{Y3}	0.025	0.163	0.155	0.951
a_{Y4}	0.010	0.075	0.074	0.943	b_{Y4}	−0.003	0.106	0.102	0.952
a_{Y5}	0.001	0.064	0.063	0.948	b_{Y5}	0.010	0.108	0.106	0.950
a_{Y6}	0.004	0.049	0.046	0.960	b_{Y6}	0.010	0.163	0.149	0.967
a_{Y7}	0.033	0.062	0.063	0.957	b_{Y7}	0.160	0.264	0.266	0.922
a_{Y8}	−0.006	0.056	0.058	0.944	b_{Y8}	0.003	0.082	0.082	0.939
a_{Y9}	0.006	0.084	0.087	0.942	b_{Y9}	0.004	0.096	0.100	0.929
a_{Y10}	−0.007	0.072	0.075	0.926	b_{Y10}	0.005	0.057	0.058	0.941
a_{Y11}	−0.001	0.064	0.064	0.947	b_{Y11}	0.011	0.082	0.084	0.936
a_{Y12}	−0.003	0.060	0.061	0.947	b_{Y12}	−0.005	0.074	0.075	0.949
a_{T1}	−0.002	0.073	0.072	0.953	b_{T1}	0.000	0.014	0.013	0.957
a_{T2}	−0.001	0.062	0.059	0.944	b_{T2}	0.000	0.012	0.013	0.939
a_{T3}	−0.002	0.062	0.062	0.953	b_{T3}	0.000	0.014	0.014	0.935
a_{T4}	0.002	0.040	0.040	0.945	b_{T4}	−0.001	0.015	0.015	0.936
a_{T5}	0.003	0.097	0.100	0.942	b_{T5}	0.000	0.012	0.012	0.946
a_{T6}	−0.002	0.062	0.064	0.935	b_{T6}	−0.001	0.012	0.012	0.942
a_{T7}	0.002	0.055	0.056	0.942	b_{T7}	0.000	0.021	0.021	0.947
a_{T8}	0.002	0.051	0.050	0.946	b_{T8}	−0.001	0.014	0.014	0.944
a_{T9}	−0.003	0.082	0.082	0.945	b_{T9}	0.000	0.014	0.014	0.945
a_{T10}	0.001	0.068	0.069	0.946	b_{T10}	−0.001	0.013	0.013	0.946
a_{T11}	0.001	0.046	0.046	0.949	b_{T11}	0.000	0.015	0.014	0.955
a_{T12}	−0.003	0.071	0.070	0.944	b_{T12}	0.000	0.013	0.013	0.936

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 10. Continued

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
b_{W1}	−0.023	0.385	0.357	0.959	b_{R1}	0.004	0.042	0.042	0.939
b_{W2}	−0.070	0.453	0.420	0.959	b_{R2}	0.000	0.066	0.064	0.946
b_{W3}	0.022	0.099	0.104	0.941	b_{R3}	0.002	0.046	0.047	0.943
b_{W4}	0.039	0.338	0.356	0.941	b_{R4}	−0.003	0.066	0.066	0.949
b_{W5}	0.044	0.247	0.254	0.951	b_{R5}	0.001	0.050	0.048	0.952
b_{W6}	0.043	0.256	0.255	0.953	b_{R6}	−0.006	0.107	0.110	0.931
b_{W7}	0.015	0.091	0.093	0.942	b_{R7}	0.002	0.040	0.040	0.947
b_{W8}	0.031	0.160	0.166	0.943	b_{R8}	0.001	0.057	0.057	0.952
b_{W9}	0.047	0.181	0.195	0.934	b_{R9}	0.002	0.044	0.044	0.950
b_{W10}	0.029	0.323	0.310	0.959	b_{R10}	0.004	0.057	0.059	0.954
b_{W11}	0.050	0.246	0.253	0.947	b_{R11}	0.004	0.069	0.071	0.945
b_{W12}	−0.003	0.053	0.052	0.960	σ_T	0.000	0.008	0.008	0.940
η_T	0.002	0.026	0.026	0.940					
η_W	0.015	0.073	0.078	0.928					
α_W	0.042	0.230	0.195	0.959					
η_R	0.002	0.032	0.033	0.934					
α_R	0.007	0.064	0.065	0.937					

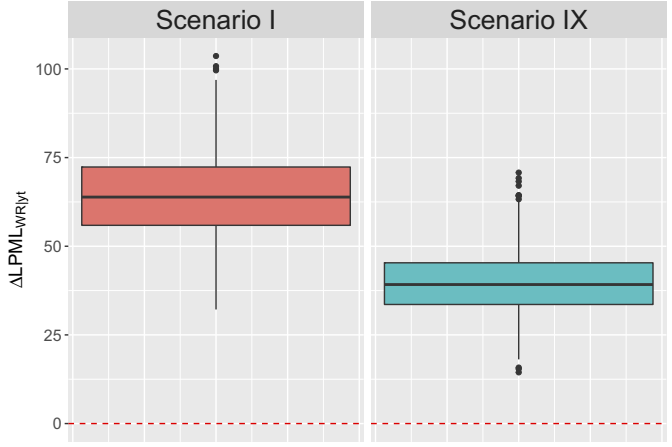


Figure 6. Boxplots of $\Delta\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ (Nonignorable–Ignorable) when the true missing data mechanisms are nonignorable (Simulation Scenario I in red) and nonignorable (Simulation Scenario IX in cyan).

greatest. Additionally, Table 15 shows that $\eta_T = -0.377$ (95% HPD = $(-0.438, -0.309)$) has a negative significant effect, indicating that subjects tend to spend less time if their latent propensities are low. $\eta_W = -0.275$ (95% HPD = $(-0.450, -0.093)$), $\alpha_W = 1.274$ (95% HPD = $(0.666, 1.909)$), $\eta_R = -0.462$ (95% HPD = $(-0.530, -0.391)$), and $\alpha_R = 1.713$ (95% HPD = $(1.552, 1.885)$) are all significantly different from 0, indicating that the missing data mechanism may be nonignorable, which is consistent with the $\text{LPML}_{\mathbf{WR}|\mathbf{yt}}$ model selection criterion in Section 6.3 and the statements in Glas and Pimentel (2008) and Rose et al. (2017). Additionally, $\sigma_T = 0.390$ (95% HPD = $(0.374, 0.404)$).

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 11. Bias, SD, RMSE, and CP of parameters in the item response model, the response time model, and the missing data models under the nonignorable missing data mechanism in Simulation Scenario IX (n = 2,000, J = 12, median sizes of ηs, 20% missing data percentage (5% due to omission + 15% due to not-reached), and the true missing data mechanism is nonignorable)

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
a_{Y1}	−0.009	0.084	0.086	0.941	b_{Y1}	0.021	0.105	0.101	0.949
a_{Y2}	−0.001	0.073	0.070	0.962	b_{Y2}	0.004	0.128	0.128	0.948
a_{Y3}	−0.006	0.067	0.064	0.950	b_{Y3}	0.050	0.231	0.209	0.955
a_{Y4}	0.021	0.105	0.105	0.939	b_{Y4}	−0.004	0.149	0.143	0.957
a_{Y5}	0.004	0.087	0.081	0.965	b_{Y5}	0.025	0.153	0.144	0.961
a_{Y6}	0.015	0.069	0.064	0.968	b_{Y6}	−0.013	0.222	0.191	0.961
a_{Y7}	0.037	0.071	0.071	0.963	b_{Y7}	0.178	0.288	0.296	0.918
a_{Y8}	−0.006	0.081	0.082	0.939	b_{Y8}	0.011	0.124	0.124	0.935
a_{Y9}	0.018	0.114	0.114	0.942	b_{Y9}	−0.004	0.133	0.128	0.946
a_{Y10}	−0.013	0.103	0.104	0.941	b_{Y10}	0.012	0.085	0.088	0.942
a_{Y11}	−0.008	0.093	0.090	0.955	b_{Y11}	0.021	0.124	0.126	0.939
a_{Y12}	−0.004	0.090	0.092	0.931	b_{Y12}	0.005	0.118	0.115	0.945
a_{T1}	0.004	0.093	0.094	0.945	b_{T1}	0.000	0.017	0.018	0.935
a_{T2}	0.000	0.085	0.088	0.936	b_{T2}	−0.001	0.017	0.017	0.944
a_{T3}	0.001	0.083	0.085	0.942	b_{T3}	−0.001	0.019	0.019	0.943
a_{T4}	0.001	0.057	0.059	0.942	b_{T4}	0.000	0.021	0.020	0.960
a_{T5}	−0.006	0.132	0.133	0.949	b_{T5}	0.000	0.016	0.015	0.955
a_{T6}	0.003	0.089	0.090	0.942	b_{T6}	−0.001	0.018	0.018	0.952
a_{T7}	0.005	0.065	0.063	0.954	b_{T7}	0.000	0.025	0.025	0.945
a_{T8}	0.002	0.073	0.072	0.947	b_{T8}	−0.002	0.021	0.021	0.945
a_{T9}	0.003	0.111	0.108	0.951	b_{T9}	−0.001	0.019	0.019	0.949
a_{T10}	−0.001	0.098	0.098	0.954	b_{T10}	−0.001	0.019	0.019	0.942
a_{T11}	0.002	0.068	0.065	0.964	b_{T11}	0.000	0.022	0.022	0.935
a_{T12}	0.002	0.106	0.105	0.950	b_{T12}	−0.001	0.020	0.019	0.956
b_{W1}	0.005	0.340	0.350	0.947	b_{R1}	0.010	0.085	0.084	0.948
b_{W2}	−0.049	0.399	0.375	0.958	b_{R2}	0.005	0.175	0.178	0.949
b_{W3}	0.019	0.095	0.098	0.939	b_{R3}	0.010	0.105	0.106	0.952
b_{W4}	0.024	0.308	0.315	0.955	b_{R4}	0.010	0.177	0.171	0.954
b_{W5}	0.031	0.228	0.235	0.954	b_{R5}	0.015	0.121	0.125	0.938
b_{W6}	0.041	0.239	0.239	0.952	b_{R6}	−0.050	0.316	0.316	0.950
b_{W7}	0.012	0.093	0.096	0.941	b_{R7}	0.013	0.067	0.067	0.951
b_{W8}	0.030	0.161	0.166	0.945	b_{R8}	0.019	0.150	0.149	0.954
b_{W9}	0.046	0.182	0.190	0.945	b_{R9}	0.014	0.098	0.102	0.940
b_{W10}	0.035	0.325	0.328	0.960	b_{R10}	0.014	0.152	0.151	0.945
b_{W11}	0.052	0.249	0.257	0.949	b_{R11}	0.018	0.189	0.189	0.948

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 11. Continued

Parameter	Bias	SD	RMSE	CP	Parameter	Bias	SD	RMSE	CP
b_{W12}	0.004	0.084	0.081	0.953	σ_T	0.000	0.008	0.008	0.940
η_T	0.002	0.036	0.037	0.936					
η_W	0.018	0.100	0.108	0.913					
α_W	0.068	0.346	0.243	0.959					
η_R	0.001	0.060	0.058	0.947					
α_R	0.022	0.119	0.124	0.936					

Table 12. The not-reached and omitted behaviors in PISA 2018 test data: the (i,j)th element represents the total number of missing answers for item j among subjects who fail to reach item i

W\Item no.	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	7	7	7	7	7	7	7	7	7	7	7	7
3	**0**	0	17	17	17	17	17	17	17	17	17	17	17
4	**0**	**0**	0	1	1	1	1	1	1	1	1	1	1
5	**1**	**1**	**28**	0	32	32	32	32	32	32	32	32	32
6	**0**	**0**	**0**	**0**	0	1	1	1	1	1	1	1	1
7	**1**	**1**	**13**	**6**	**17**	0	34	34	34	34	34	34	34
8	**0**	**0**	**5**	**6**	**8**	**3**	0	12	12	12	12	12	12
9	**6**	**1**	**31**	**22**	**30**	**7**	**22**	0	47	47	47	47	47
10	**3**	**5**	**27**	**22**	**32**	**8**	**14**	**8**	0	53	53	53	53
11	**2**	**0**	**29**	**16**	**35**	**5**	**14**	**2**	**16**	0	63	63	63
12	**3**	**0**	**23**	**11**	**25**	**4**	**9**	**4**	**6**	**17**	0	58	58
13	**22**	**9**	**413**	**251**	**497**	**58**	**117**	**43**	**83**	**200**	**332**	0	1,417
(Total													
omitted	**(38)**	**(17)**	**(569)**	**(334)**	**(644)**	**(85)**	**(176)**	**(57)**	**(105)**	**(217)**	**(332)**	**(0)**	
missingness)													
Total (R)	38	24	593	359	701	143	268	161	256	421	599	325	

According to Table 16, the correlation of the latent propensities between the item response and the not-reached behavior, measured by $\eta_W sin(\alpha_W) = -0.252$ (95% HPD = $(-0.423, -0.081)$) is significant, indicating that subjects with better latent abilities tend not to withdraw the test. The correlation of the latent propensities between the response time and the not-reached behavior, measured by $\eta_W(\eta_T \sin(\alpha_W) + \sqrt{1-(\eta_T)^2}\cos(\alpha_W)) = 0.018$ (95% HPD = $(-0.126, 0.151)$) is insignificant. We then calculate the correlation of latent propensities between the item response and the omitted behavior, measured by $\eta_R sin(\alpha_R) = -0.456$ (95% HPD = $(-0.524, -0.382)$), indicating that subjects with better latent abilities tend not to omit items. We further compute the correlation of latent propensities between the response time and the omitted behavior, measured by $\eta_R(\eta_T \sin(\alpha_R) + \sqrt{1-(\eta_T)^2}\cos(\alpha_R)) = 0.232$ (95% HPD = $(0.164, 0.286)$). The positive signficant value indicates that subjects with longer response times tend to omit items.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 13. The total number of observed answers (n_{obs}), the number and percentage of correct answers, the median, the first quartile (Q_1), and the third quartile (Q_3) of response times (in minutes) for each item

Item no.	1	2	3	4	5	6	7	8	9	10	11	12
n_{obs}	1,704	1,718	1,149	1,383	1,041	1,599	1,474	1,581	1,486	1,321	1,143	1,417
Correct	1,129	651	554	714	197	514	194	441	67	148	31	198
%	66.26	37.89	48.22	51.63	18.92	32.15	13.16	27.89	4.51	11.20	2.71	13.97
Median	8.64	10.07	29.54	14.38	19.65	14.38	27.46	9.56	16.79	27.22	13.49	15.69
Q_1	6.22	7.39	17.65	8.63	9.09	9.38	17.93	5.98	12.23	19.05	8.46	10.24
Q_3	12.58	14.86	47.32	21.40	35.71	20.03	38.95	14.01	24.12	37.36	22.37	22.28

Table 14. Mean, SD, and 95% HPD of the difficulty parameters in the item response model under the nonignorable missing data mechanism

Parameter	Mean	SD	95% HPD
$a_{Y1}b_{Y1}$	−0.759	0.065	(−0.883, −0.642)
$a_{Y2}b_{Y2}$	0.631	0.063	(0.501, 0.748)
$a_{Y3}b_{Y3}$	0.612	0.117	(0.390, 0.836)
$a_{Y4}b_{Y4}$	0.191	0.080	(0.022, 0.337)
$a_{Y5}b_{Y5}$	2.007	0.122	(1.797, 2.266)
$a_{Y6}b_{Y6}$	1.001	0.075	(0.846, 1.146)
$a_{Y7}b_{Y7}$	3.315	0.243	(2.879, 3.775)
$a_{Y8}b_{Y8}$	0.977	0.060	(0.868, 1.108)
$a_{Y9}b_{Y9}$	5.062	0.414	(4.400, 5.958)
$a_{Y10}b_{Y10}$	3.640	0.285	(3.143, 4.226)
$a_{Y11}b_{Y11}$	5.366	0.455	(4.629, 6.300)
$a_{Y12}b_{Y12}$	2.389	0.129	(2.155, 2.651)

We also present the posterior summaries under the ignorable missing data mechanism in Table 17, some of which are quite different from the posterior summaries under the nonignorable missing data mechanism in Table 15. For example, $b_{Y4} = 0.120$ is significant under the nonignorable missing data mechanism (95% HPD = $(0.015, 0.211)$ but is not significant under the ignorable missing data mechanism $b_{Y4} = 0.035$ (95% HPD = $(-0.072, 0.126)$).

6.3. Model comparison

We first use the LPML decomposition technique developed in Section 4.3 to guide the selection of the missing data mechanism. The $\text{LPML}_{\mathbf{WR|yt}}$ (−6943.24) under the nonignorable missing data mechanism is greater than the $\text{LPML}_{\mathbf{WR}}$ (−7026.59) under the ignorable missing data mechanism, indicating that the nonignorable missing data mechanism is more favorable for this dataset. The posterior summaries under the nonignorable missing data mechanism are given in Tables 14 and 15.

To further evaluate the significance of the joint model for item response ($\mathbf{y}$) and response time ($\mathbf{t}$) in identifying the missing data mechanism, we compare it with the single outcome model for $\mathbf{y}$ in (19) and the single outcome model for $\mathbf{t}$ in (20) under the nonignorable missing data mechanism, respectively. The $\text{LPML}_{\mathbf{WR|y}}$ (−6957.56) for the $\mathbf{y}$-only model and the $\text{LPML}_{\mathbf{WR|t}}$ (−6999.81)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 15. Mean, SD, and 95% HPD of discrimination parameters and difficulty parameters in the item response model, item time intensity parameters and discrimination parameters in the response time model, difficulty parameters for the not-reached behavior, difficulty parameters for the omitted behavior, and the formulation parameters under the nonignorable missing data mechanism

Parameter	Mean	SD	95% HPD	Parameter	Mean	SD	95% HPD
a_{Y1}	0.871	0.082	(0.725, 1.029)	b_{Y1}	−0.878	0.096	(−1.074, −0.700)
a_{Y2}	1.033	0.086	(0.883, 1.222)	b_{Y2}	0.614	0.068	(0.478, 0.743)
a_{Y3}	2.138	0.194	(1.759, 2.533)	b_{Y3}	0.286	0.048	(0.193, 0.377)
a_{Y4}	1.587	0.143	(1.315, 1.881)	b_{Y4}	0.120	0.049	(0.015, 0.211)
a_{Y5}	1.101	0.124	(0.858, 1.364)	b_{Y5}	1.837	0.158	(1.582, 2.216)
a_{Y6}	1.139	0.091	(0.967, 1.316)	b_{Y6}	0.881	0.071	(0.753, 1.023)
a_{Y7}	2.171	0.226	(1.752, 2.629)	b_{Y7}	1.533	0.079	(1.391, 1.709)
a_{Y8}	0.346	0.057	(0.239, 0.466)	b_{Y8}	2.898	0.482	(2.135, 4.095)
a_{Y9}	2.268	0.272	(1.818, 2.880)	b_{Y9}	2.244	0.127	(2.020, 2.500)
a_{Y10}	2.171	0.266	(1.745, 2.762)	b_{Y10}	1.687	0.102	(1.489, 1.900)
a_{Y11}	1.933	0.272	(1.496, 2.485)	b_{Y11}	2.803	0.228	(2.426, 3.324)
a_{Y12}	1.274	0.129	(1.025, 1.505)	b_{Y12}	1.888	0.149	(1.645, 2.195)
a_{T1}	4.276	0.170	(3.954, 4.631)	b_{T1}	2.180	0.015	(2.148, 2.209)
a_{T2}	3.738	0.147	(3.456, 4.011)	b_{T2}	2.361	0.016	(2.329, 2.392)
a_{T3}	0.865	0.037	(0.797, 0.942)	b_{T3}	3.125	0.032	(3.066, 3.186)
a_{T4}	0.659	0.026	(0.611, 0.709)	b_{T4}	2.284	0.036	(2.215, 2.362)
a_{T5}	0.781	0.036	(0.714, 0.858)	b_{T5}	2.762	0.036	(2.696, 2.832)
a_{T6}	1.914	0.072	(1.789, 2.062)	b_{T6}	2.490	0.021	(2.448, 2.528)
a_{T7}	1.101	0.041	(1.030, 1.186)	b_{T7}	3.088	0.029	(3.036, 3.147)
a_{T8}	2.688	0.107	(2.470, 2.892)	b_{T8}	2.123	0.018	(2.087, 2.159)
a_{T9}	5.427	0.238	(4.976, 5.919)	b_{T9}	2.806	0.016	(2.776, 2.836)
a_{T10}	5.080	0.235	(4.662, 5.570)	b_{T10}	3.209	0.016	(3.179, 3.242)
a_{T11}	2.866	0.132	(2.633, 3.126)	b_{T11}	2.618	0.020	(2.578, 2.657)
a_{T12}	3.343	0.141	(3.075, 3.623)	b_{T12}	2.698	0.018	(2.665, 2.732)
b_{W1}	8.491	1.159	(6.634, 11.185)	b_{R1}	4.438	0.169	(4.136, 4.762)
b_{W2}	6.073	0.404	(5.414, 6.956)	b_{R2}	5.236	0.249	(4.769, 5.776)
b_{W3}	5.142	0.255	(4.664, 5.650)	b_{R3}	0.872	0.063	(0.749, 1.008)
b_{W4}	7.633	0.871	(6.285, 9.623)	b_{R4}	1.728	0.074	(1.588, 1.868)
b_{W5}	4.449	0.184	(4.121, 4.874)	b_{R5}	0.584	0.063	(0.457, 0.718)
b_{W6}	7.624	0.854	(6.344, 9.547)	b_{R6}	3.476	0.119	(3.250, 3.717)
b_{W7}	4.333	0.177	(3.991, 4.681)	b_{R7}	2.576	0.092	(2.391, 2.743)
b_{W8}	5.376	0.306	(4.817, 6.052)	b_{R8}	3.848	0.144	(3.570, 4.135)
b_{W9}	3.914	0.148	(3.618, 4.199)	b_{R9}	3.080	0.111	(2.843, 3.298)
b_{W10}	3.710	0.152	(3.437, 4.008)	b_{R10}	2.069	0.090	(1.898, 2.245)
b_{W11}	3.464	0.138	(3.205, 3.745)	b_{R11}	1.371	0.074	(1.228, 1.518)

(Continued)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 15. Continued

Parameter	Mean	SD	95% HPD	Parameter	Mean	SD	95% HPD
b_{W12}	3.429	0.137	(3.153, 3.721)	σ_T	0.390	0.008	(0.374, 0.404)
η_T	−0.377	0.034	(−0.438, −0.309)				
η_W	−0.275	0.093	(−0.450, −0.093)				
α_W	1.274	0.299	(0.666, 1.909)				
η_R	−0.462	0.037	(−0.530, −0.391)				
α_R	1.713	0.085	(1.552, 1.885)				

Table 16. Mean, SD, and 95% HPD of computed parameters under the nonignorable missing data mechanism

Parameter	Mean	SD	95% HPD
$\eta_W \sin(\alpha_W)$	−0.252	0.088	(−0.423, −0.081)
$\eta_W(\eta_T \sin(\alpha_W) + \sqrt{1-(\eta_T)^2}\cos(\alpha_W))$	0.018	0.070	(−0.126, 0.151)
$\eta_R \sin(\alpha_R)$	−0.456	0.039	(−0.524, −0.382)
$\eta_R(\eta_T \sin(\alpha_R) + \sqrt{1-(\eta_T)^2}\cos(\alpha_R))$	0.232	0.031	(0.164, 0.286)

for the **t**-only model are smaller than that for the joint model, indicating that the joint model provides better identification of the missing data mechanism compared to the two single outcome models.

We further compare our methods with MICE with 20 imputations. We apply the same item response and response time models in (1) and (2) for each imputed data set for fair comparison. Table 18 shows the final posterior summaries obtained after pooling. Most results are similar except for b_{Y4}, which is not significant under MICE but is significant under the proposed nonignorable missing data models. The computation time under MICE (274.87 minutes) is substantially longer than that of the proposed model (49.00 minutes). Additionally, the omission and not-reached behaviors are not studied in the MICE approach. We also compare our methods with the RTprocess model in Table 19. For fair comparison, we assume the same hierarchical framework and reparameterization technique for θ and θ_T, which guarantees that the covariance matrix is positive definite. Similarly to MICE, b_{Y4} is not significant under the RTprocess model but is significant under the proposed nonignorable missing data models. η_T and σ_T are similar to those obtained in the proposed model. However, similar to MICE, the omission and not-reached behaviors are not modeled directly.

Finally, we compare the estimations and inferences of the ability parameters across different models. The stacked bar charts in Figure 7 show the counts of significant and insignificant ability parameter estimates $\hat{\theta}$, grouped by the number of missing items. The bars represent negative significant estimates with 95% HPD intervals that are entirely below zero in blue, positive significant estimates with 95% HPD intervals that are entirely above zero in green, and insignificant estimates with 95% HPD intervals including zero in red. We see from Figure 7 that there are lower (higher) proportions of "blue" and higher (lower) proportions of "green" among the bars for small (large) numbers of missing items under the proposed nonignorable model than the other three methods. The results under the nonignorable model are more desirable and intuitively appealing since the subjects with higher (lower) abilities are more likely to have fewer (more) missing items. Table 20 presents the minimum (Min), first quartile (Q1), median (Q2), third quartile (Q3), and maximum (Max) of the posterior means of the latent traits across different models. The proposed nonignorable model has the largest values of IQR (Q3–Q1) and range (Max–Min). Thus, the nonignorable model can distinguish the subjects' abilities more effectively.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 17. Mean, SD, and 95% HPD of discrimination parameters and difficulty parameters in the item response model, item time intensity parameters and discrimination parameters in the response time model, difficulty parameters for the not-reached behavior, and difficulty parameters for the omitted behavior under the ignorable missing data mechanism

Parameter	Mean	SD	95% HPD	Parameter	Mean	SD	95% HPD
a_{Y1}	0.978	0.094	(0.800, 1.183)	b_{Y1}	−0.813	0.085	(−0.991, −0.655)
a_{Y2}	1.160	0.098	(0.974, 1.361)	b_{Y2}	0.552	0.060	(0.432, 0.674)
a_{Y3}	2.098	0.204	(1.714, 2.518)	b_{Y3}	0.173	0.047	(0.084, 0.264)
a_{Y4}	1.471	0.131	(1.243, 1.729)	b_{Y4}	0.035	0.051	(−0.072, 0.126)
a_{Y5}	1.121	0.133	(0.871, 1.407)	b_{Y5}	1.720	0.157	(1.436, 2.029)
a_{Y6}	1.069	0.091	(0.898, 1.251)	b_{Y6}	0.873	0.077	(0.728, 1.049)
a_{Y7}	2.137	0.214	(1.753, 2.583)	b_{Y7}	1.491	0.084	(1.343, 1.673)
a_{Y8}	0.314	0.059	(0.206, 0.427)	b_{Y8}	3.179	0.639	(2.223, 4.624)
a_{Y9}	2.163	0.266	(1.693, 2.698)	b_{Y9}	2.247	0.149	(1.992, 2.568)
a_{Y10}	2.067	0.240	(1.685, 2.658)	b_{Y10}	1.655	0.095	(1.485, 1.841)
a_{Y11}	1.742	0.287	(1.272, 2.318)	b_{Y11}	2.893	0.276	(2.418, 3.514)
a_{Y12}	1.225	0.125	(0.998, 1.480)	b_{Y12}	1.882	0.147	(1.630, 2.197)
a_{T1}	4.394	0.169	(4.074, 4.740)	b_{T1}	2.183	0.016	(2.153, 2.214)
a_{T2}	3.779	0.161	(3.484, 4.098)	b_{T2}	2.361	0.015	(2.333, 2.391)
a_{T3}	0.864	0.039	(0.791, 0.940)	b_{T3}	3.134	0.034	(3.070, 3.201)
a_{T4}	0.661	0.026	(0.611, 0.713)	b_{T4}	2.290	0.033	(2.223, 2.354)
a_{T5}	0.772	0.034	(0.706, 0.836)	b_{T5}	2.774	0.035	(2.704, 2.839)
a_{T6}	1.914	0.074	(1.778, 2.074)	b_{T6}	2.493	0.020	(2.452, 2.531)
a_{T7}	1.101	0.042	(1.022, 1.184)	b_{T7}	3.091	0.027	(3.038, 3.146)
a_{T8}	2.670	0.103	(2.477, 2.863)	b_{T8}	2.126	0.018	(2.093, 2.161)
a_{T9}	5.523	0.252	(5.030, 6.004)	b_{T9}	2.810	0.015	(2.781, 2.839)
a_{T10}	4.972	0.224	(4.554, 5.411)	b_{T10}	3.216	0.016	(3.185, 3.247)
a_{T11}	2.829	0.129	(2.593, 3.090)	b_{T11}	2.626	0.019	(2.588, 2.665)
a_{T12}	3.283	0.136	(3.030, 3.576)	b_{T12}	2.700	0.018	(2.664, 2.733)
b_{W1}	8.479	1.186	(6.671, 11.281)	b_{R1}	4.418	0.170	(4.103, 4.755)
b_{W2}	6.064	0.379	(5.343, 6.849)	b_{R2}	5.238	0.236	(4.820, 5.720)
b_{W3}	5.096	0.249	(4.657, 5.610)	b_{R3}	0.868	0.062	(0.743, 0.992)
b_{W4}	7.649	0.812	(6.283, 9.395)	b_{R4}	1.727	0.076	(1.578, 1.878)
b_{W5}	4.427	0.181	(4.093, 4.806)	b_{R5}	0.581	0.061	(0.463, 0.705)
b_{W6}	7.552	0.819	(6.251, 9.287)	b_{R6}	3.481	0.126	(3.249, 3.734)
b_{W7}	4.340	0.190	(3.987, 4.730)	b_{R7}	2.579	0.090	(2.410, 2.763)
b_{W8}	5.348	0.279	(4.848, 5.918)	b_{R8}	3.853	0.150	(3.593, 4.157)
b_{W9}	3.904	0.155	(3.611, 4.200)	b_{R9}	3.073	0.112	(2.854, 3.302)
b_{W10}	3.725	0.153	(3.455, 4.030)	b_{R10}	2.064	0.084	(1.909, 2.231)
b_{W11}	3.454	0.134	(3.197, 3.709)	b_{R11}	1.369	0.073	(1.224, 1.509)
b_{W12}	3.433	0.140	(3.181, 3.718)	σ_T	0.391	0.009	(0.373, 0.407)

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 18. Mean, SD, and 95% HPD of discrimination parameters and difficulty parameters in the item response model and item time intensity parameters and discrimination parameters in the response time model under MICE

Parameter	Mean	SD	95% HPD	Parameter	Mean	SD	95% HPD
a_{Y1}	0.941	0.093	(0.764, 1.131)	b_{Y1}	−0.834	0.090	(−1.025, −0.672)
a_{Y2}	1.151	0.096	(0.974, 1.348)	b_{Y2}	0.553	0.063	(0.437, 0.682)
a_{Y3}	2.056	0.199	(1.703, 2.475)	b_{Y3}	0.191	0.046	(0.101, 0.282)
a_{Y4}	1.522	0.128	(1.283, 1.782)	b_{Y4}	0.046	0.049	(−0.049, 0.143)
a_{Y5}	0.986	0.130	(0.741, 1.246)	b_{Y5}	1.846	0.192	(1.528, 2.281)
a_{Y6}	1.137	0.099	(0.948, 1.339)	b_{Y6}	0.843	0.078	(0.700, 1.003)
a_{Y7}	2.111	0.211	(1.723, 2.547)	b_{Y7}	1.484	0.082	(1.334, 1.656)
a_{Y8}	0.333	0.060	(0.221, 0.456)	b_{Y8}	2.990	0.561	(2.129, 4.317)
a_{Y9}	2.133	0.289	(1.617, 2.754)	b_{Y9}	2.233	0.153	(1.971, 2.572)
a_{Y10}	2.094	0.230	(1.682, 2.580)	b_{Y10}	1.626	0.096	(1.456, 1.826)
a_{Y11}	1.608	0.267	(1.144, 2.199)	b_{Y11}	2.902	0.294	(2.416, 3.578)
a_{Y12}	1.291	0.134	(1.042, 1.568)	b_{Y12}	1.818	0.140	(1.570, 2.117)
a_{T1}	4.158	0.167	(3.836, 4.494)	b_{T1}	2.184	0.015	(2.155, 2.213)
a_{T2}	3.675	0.144	(3.405, 3.960)	b_{T2}	2.362	0.016	(2.331, 2.392)
a_{T3}	0.820	0.049	(0.728, 0.924)	b_{T3}	3.121	0.034	(3.052, 3.186)
a_{T4}	0.637	0.030	(0.580, 0.694)	b_{T4}	2.287	0.035	(2.217, 2.355)
a_{T5}	0.741	0.036	(0.671, 0.812)	b_{T5}	2.725	0.035	(2.655, 2.792)
a_{T6}	1.928	0.079	(1.774, 2.088)	b_{T6}	2.493	0.021	(2.452, 2.533)
a_{T7}	1.085	0.046	(0.995, 1.175)	b_{T7}	3.086	0.027	(3.034, 3.139)
a_{T8}	2.688	0.106	(2.485, 2.898)	b_{T8}	2.122	0.018	(2.087, 2.157)
a_{T9}	5.348	0.240	(4.893, 5.819)	b_{T9}	2.807	0.015	(2.779, 2.836)
a_{T10}	4.955	0.254	(4.467, 5.467)	b_{T10}	3.211	0.016	(3.180, 3.243)
a_{T11}	2.772	0.130	(2.527, 3.035)	b_{T11}	2.620	0.020	(2.582, 2.659)
a_{T12}	3.338	0.139	(3.074, 3.618)	b_{T12}	2.700	0.017	(2.666, 2.734)
σ_T	0.392	0.009	(0.376, 0.409)				
η_T	−0.369	0.033	(−0.432, −0.304)				

Figure 8 shows the differences between ability estimates under the proposed nonignorable model and the competing models as a function of the response time propensity estimates under the proposed nonignorable model. Data points with darker color correspond to subjects with more missing items. Among all the competing approaches, MICE provides the most different results. The ability estimates under the proposed nonignorable model tend to be smaller than those under the proposed ignorable model when the number of missing items is high and when the response time propensity estimate is large. This suggests that the missing data process influences the analysis of outcome models, indicating that the missing data mechanism may be nonignorable. The ability estimates under the proposed nonignorable model tend to be smaller than those under the RTprocess model for subjects with more missing items. The results are consistent with intuition that subjects with higher (lower) abilities are more likely to have fewer (more) missing items.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Table 19. Mean, SD, and 95% HPD of discrimination parameters and difficulty parameters in the item response model and item time intensity parameters and discrimination parameters in the response time model under the response time process model

Parameter	Mean	SD	95% HPD	Parameter	Mean	SD	95% HPD
a_{Y1}	0.954	0.092	(0.786, 1.143)	b_{Y1}	−0.838	0.086	(−1.004, −0.673)
a_{Y2}	1.168	0.094	(1.002, 1.369)	b_{Y2}	0.537	0.063	(0.413, 0.664)
a_{Y3}	2.073	0.189	(1.717, 2.461)	b_{Y3}	0.167	0.046	(0.084, 0.256)
a_{Y4}	1.519	0.128	(1.287, 1.809)	b_{Y4}	0.043	0.048	(−0.054, 0.130)
a_{Y5}	1.125	0.122	(0.894, 1.383)	b_{Y5}	1.685	0.151	(1.405, 1.998)
a_{Y6}	1.133	0.103	(0.941, 1.335)	b_{Y6}	0.830	0.075	(0.689, 0.982)
a_{Y7}	2.211	0.226	(1.813, 2.663)	b_{Y7}	1.440	0.083	(1.288, 1.610)
a_{Y8}	0.378	0.059	(0.271, 0.495)	b_{Y8}	2.589	0.403	(1.939, 3.463)
a_{Y9}	2.316	0.303	(1.821, 2.925)	b_{Y9}	2.136	0.148	(1.867, 2.451)
a_{Y10}	2.240	0.256	(1.752, 2.787)	b_{Y10}	1.568	0.095	(1.399, 1.749)
a_{Y11}	1.980	0.271	(1.503, 2.582)	b_{Y11}	2.633	0.211	(2.271, 3.107)
a_{Y12}	1.333	0.132	(1.087, 1.604)	b_{Y12}	1.763	0.133	(1.520, 2.057)
a_{T1}	4.236	0.165	(3.908, 4.555)	b_{T1}	2.186	0.016	(2.155, 2.218)
a_{T2}	3.660	0.142	(3.395, 3.948)	b_{T2}	2.365	0.016	(2.334, 2.396)
a_{T3}	0.872	0.035	(0.814, 0.942)	b_{T3}	3.352	0.031	(3.285, 3.410)
a_{T4}	0.668	0.026	(0.615, 0.719)	b_{T4}	2.362	0.034	(2.294, 2.428)
a_{T5}	0.797	0.036	(0.731, 0.875)	b_{T5}	2.974	0.034	(2.911, 3.036)
a_{T6}	1.921	0.069	(1.787, 2.066)	b_{T6}	2.494	0.019	(2.456, 2.533)
a_{T7}	1.098	0.044	(1.010, 1.188)	b_{T7}	3.151	0.027	(3.095, 3.201)
a_{T8}	2.645	0.102	(2.449, 2.840)	b_{T8}	2.132	0.019	(2.094, 2.168)
a_{T9}	5.318	0.228	(4.888, 5.774)	b_{T9}	2.833	0.014	(2.806, 2.859)
a_{T10}	4.960	0.220	(4.547, 5.441)	b_{T10}	3.238	0.015	(3.208, 3.268)
a_{T11}	2.765	0.118	(2.541, 3.002)	b_{T11}	2.690	0.020	(2.653, 2.727)
a_{T12}	3.304	0.139	(3.029, 3.573)	b_{T12}	2.700	0.017	(2.671, 2.732)
σ_T	0.395	0.008	(0.379, 0.413)				
η_T	−0.362	0.033	(−0.420, −0.291)				

Table 20. Minimum, first quartile, median, third quartile, and maximum of the posterior means of the ability parameters

Method	Min	Q1	Q2	Q3	Max
Nonignorable	−2.602	−0.651	−0.070	0.607	2.871
Ignorable	−1.347	−0.636	−0.080	0.511	2.884
MICE	−2.171	−0.614	−0.066	0.523	2.759
RTprocess	−2.176	−0.657	−0.057	0.586	2.842

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

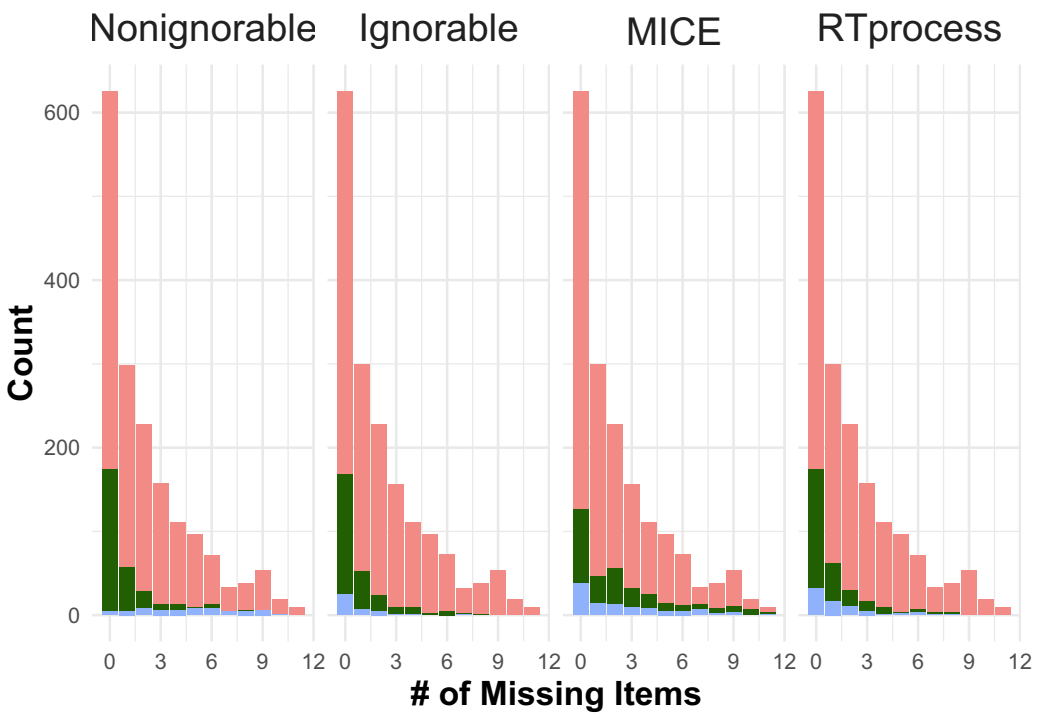


Figure 7. Stacked bar charts showing the number of significant and insignificant ability parameter estimates across different methods, grouped by the number of missing items. Blue bars correspond to the negative significant $\hat{\theta}$, green bars correspond to the positive significant $\hat{\theta}$, and red bars correspond to the insignificant $\hat{\theta}$.

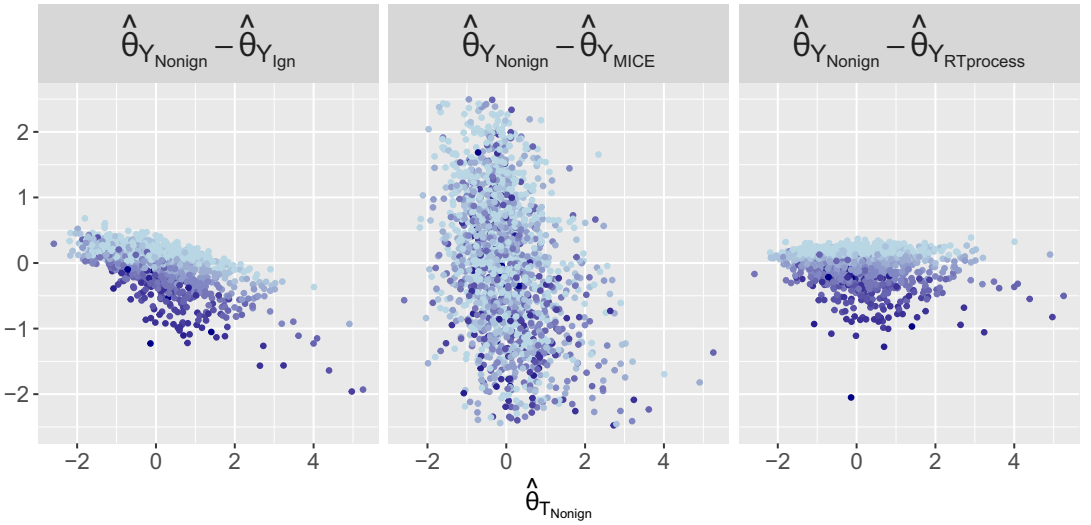


Figure 8. Difference in the ability estimates between the proposed nonignorable model and the competing models plotted against the response time propensity estimates under the proposed nonignorable model. The number of missing values is given by the data points' color, with darker colors denoting a higher number of missingness.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

7. Discussion

In this article, a Bayesian hierarchical IRT-based model is developed to jointly model the item response, response time, and not-reached and omitted behaviors. A missing data model is built via a combination of a sequential multinomial model for the not-reached behavior and a conditional model for the omitted behavior given the not-reached item. A reparameterization technique is introduced for the joint model of the latent propensity parameters. This allows us to investigate the conditional independence properties of the latent propensity parameters, measure the extent to which the missing data mechanism is nonignorable, and facilitate the specification of prior distributions that results in a more efficient Bayesian computation. Furthermore, a model selection criterion is developed to guide the selection of the missing data mechanism. It can be further used to quantify the importance of modeling item response and response time jointly over individually in identifying the missing data mechanism. The simulation studies show that the proposed approach is able to provide accurate estimations and inferences of the parameters of interest, and the model selection criterion also performs well in selecting the optimal missing data mechanism. The proposed method is further applied to the math exam data from 2018 PISA Test and our model selection criterion indicates that the missingness is more likely nonignorable. The findings from the proposed model regarding item response, response time, and omitted and not-reached behaviors align well with what we observe in the real data.

Our proposed approach allows for directly and jointly modeling of the omitted and the not-reached behaviors, without restricting missingness to time constraints, and is thus more flexible. Our definitions for ignorable and nonignorable missing data mechanisms in Remark 3.2 are more general than the definitions for latent ignorable and nonignorable missing data mechanisms in Robitzsch (2021), which is provided specifically for the shared-parameter model. We assume the missing responses depend only on the latent variables satisfying the second conditions of nonignorability in Pohl and Becker (2020). This assumption can be relaxed by incorporating covariates such as missing outcomes in the missing data model (Pohl & Becker, 2020; Robitzsch, 2023). Last but not least, our data-driven missing data model allows for nonignorable missing data mechanism and also ignorable missing data mechanism as a special case.

The current study does not distinguish aberrant behaviors (i.e., rapid guessing and cheating) (Lu et al., 2023) from the solution behavior for response and response time models. How to naturally incorporate these behaviors with the not-reached and omitted behaviors under the current model framework remains to be studied. We currently consider the Rasch model for the missing data mechanism due to identifiability issues. Other more complicated models may be considered provided with enough items and subjects or if we can incorporate techniques for borrowing information from "adjacent" parameters, such as the mixture of finite mixture (MFM) prior for parameter clustering (Hu et al., 2023; Miller & Harrison, 2018). We currently do not differentiate between different types of omitted behavior (i.e., intentional, accidental, or time-limited) and not-reached behavior (due to overall time limits or early quitting). The first avenue remains a fascinating direction for future work, while the second can be readily accomplished by redefining $\mathbf{W}$ and adjusting the response time model. Additionally, the decomposition of the LPML model selection criterion can also be modified for selecting the response models, which is left for future studies.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2026.10121.

Data availability statement. The PISA 2018 data is publicly available at https://www.oecd.org/en/data/datasets/pisa-2018-database.html.

Acknowledgements. The authors would like to thank the Editor, the Associate Editor, and the three anonymous referees for their insightful comments and suggestions, which led to an improved version of the article.

Funding statement. This research received no specific grant funding from any funding agency, commercial, or not-for-profit sectors.

Competing interests. The authors declare none.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

References

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 443–459.

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, *26*, 403–417.

Debeer, D., Janssen, R., & De Boeck, P. (2017). Modeling skipped and not-reached items using IRTrees. *Journal of Educational Measurement*, *54*(3), 333–363.

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. Springer.

Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, *74*(365), 153–160.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and draper). *Bayesian Analysis*, *1*(3), 515–534.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Glas, C. A., & Pimentel, J. L. (2008). Modeling nonignorable missing data in speeded tests. *Educational and Psychological Measurement*, *68*(6), 907–922.

Glas, C. A., Pimentel, J., & Lamers, S. (2015). Nonignorable data in IRT models: Polytomous responses and response propensity models with covariates. *Psychological Test and Assessment Modeling*, *57*(4), 523–541.

Guo, J. (2019). An IRT-based model for omitted and not-reached items. Preprint, arXiv:1904.03767.

Guo, J., Xu, X., Fang, G., Ying, Z., & Zhang, S. (2025). Jointly modeling responses and omitted items by a competing risk model: A survival analysis approach. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12382

Guo, J., Xu, X., Ying, Z., & Zhang, S. (2022). Modeling not-reached items in timed tests: A response time censoring approach. *Psychometrika*, *87*(3), 835–867.

Guo, J., Xu, X., & Zhang, S. (2025). Jointly modeling omitted and not-reached items in time-limit tests: A survival analysis approach. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/107699862513808

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. *2*). Sage.

Harwell, M. R., & Baker, F. B. (1991). The use of prior distributions in marginalized Bayesian item parameter estimation: A didactic. *Applied Psychological Measurement*, *15*(4), 375–389.

Holman, R., & Glas, C. A. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology*, *58*(1), 1–17.

Hu, G., Xue, Y., & Ma, Z. (2023). Bayesian clustered coefficients regression with auxiliary covariates assistant random effects. *Statistical Modelling*, *23*(3), 273–293.

Ibrahim, J. G., Chen, M.-H., & Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, *88*(2), 551–564.

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., & Herring, A. H. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, *100*(469), 332–346.

Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2001). *Bayesian survival analysis* (Vol. *2*). Springer.

Köhler, C., Pohl, S., & Carstensen, C. H. (2015). Taking the missing propensity into account when estimating competence scores: Evaluation of item response theory models for nonignorable omissions. *Educational and Psychological Measurement*, *75*(5), 850–874.

List, M. K., Köller, O., & Nagy, G. (2019). A semiparametric approach for modeling not-reached items. *Educational and Psychological Measurement*, *79*(1), 170–199.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Liu, F., Chen, M.-H., Wang, X., & Hancock, R. (2025). Decomposition of WAIC for assessing the information gain with application to educational testing. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12383

Lu, J., & Wang, C. (2020). A response time process model for not-reached and omitted items. *Journal of Educational Measurement*, *57*(4), 584–620.

Lu, J., Wang, C., & Shi, N. (2023). A mixture response time process model for aberrant behaviors and item nonresponses. *Multivariate Behavioral Research*, *58*(1), 71–89.

Miller, J. W., & Harrison, M. T. (2018). Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, *113*(521), 340–356.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, *51*, 177–195.

Mislevy, R. J., & Wu, P.-K. (1988). Inferring examinee ability when some item responses are missing. *ETS Research Report Series*, *1988*(2), i–75.

O'Muircheartaigh, C., & Moustaki, I. (1999). Symmetric pattern models: A latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*(2), 177–194.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*(2), 146–178.

Pohl, S., & Becker, B. (2020). Performance of missing data approaches under nonignorable missing data conditions. *Methodology*, *16*(2), 147–165.

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press

Pohl, S., Ulitzsch, E., & von Davier, M. (2019). Using response times to model not-reached items due to time limits. *Psychometrika*, *84*(3), 892–920.
Rabe-Hesketh, S., & Skrondal, A. (2023). Ignoring non-ignorable missingness. *Psychometrika*, *88*(1), 31–50.
Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. ERIC.
Robitzsch, A. (2021). On the treatment of missing item responses in educational large-scale assessment data: An illustrative simulation study and a case study using PISA 2018 mathematics data. *European Journal of Investigation in Health, Psychology and Education*, *11*(4), 1653–1687.
Robitzsch, A. (2023). Nonignorable consequences of (partially) ignoring missing item responses: Students omit (constructed response) items due to a lack of knowledge. *Knowledge*, *3*(2), 215–231.
Rose, N. (2013). *Item nonresponses in educational and psychological measurement*. [Doctoral Dissertation, Friedrich-Schiller-University, Jena, Germany]. Retrieved from https://www.db-thueringen.de/receive/dbt_mods_00022476.
Rose, N., von Davier, M.,& Nagengast, B. (2015). Commonalities and differences in IRT-based methods for nonignorable item nonresponses. *Psychological Test and Assessment Modeling*, *57*(4), 472.
Rose, N., von Davier, M., & Nagengast, B. (2017). Modeling omitted and not-reached items in IRT models. *Psychometrika*, *82*(3), 795–819.
Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.
Sun, J., Liu, Y., Wang, X., & Chen, M.-H. (2025). Bayesian variable selection in dynamic item response theory models. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/10769986251314527
Ulitzsch, E., Alfers, T., Lu, J., Wang, C., & Pohl, S. (2025). Using response times for improving the understanding and handling of missing responses in low-stakes assessments. In L. Khorramdel, M. von Davier, & K. Yamamoto (Ed.), *Innovative digital-based international large-scale assessments: Foundations, methodologies, and quality assurance*. Springer.
Ulitzsch, E., von Davier, M., & Pohl, S. (2020a). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non-response. *British Journal of Mathematical and Statistical Psychology*, *73*, 83–112.
Ulitzsch, E., von Davier, M., & Pohl, S. (2020b). A multiprocess item response model for not-reached items due to time limits and quitting. *Educational and Psychological Measurement*, *80*(3), 522–547.
Ulitzsch, E., von Davier, M., & Pohl, S. (2020c). Using response times for joint modeling of response and omission behavior. *Multivariate Behavioral Research*, *55*(3), 425–453.
Ulitzsch, E., Zhang, S., & Pohl, S. (2024). A model-based approach to the disentanglement and differential treatment of engaged and disengaged item omissions. *Multivariate Behavioral Research*, *59*(3), 599–619.
Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*, 1–67.
Van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181–204.
Van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*(3), 287–308.
Wu, J., Chen, M.-H., Schifano, E. D., Ibrahim, J. G., & Fisher, J. D. (2019). A new Bayesian joint model for longitudinal count data with many zeros, intermittent missingness, and dropout with applications to HIV prevention trials. *Statistics in Medicine*, *38*(30), 5565–5586.
Yuan, K.-H., Jamshidian, M., & Kano, Y. (2018). Missing data mechanisms and homogeneity of means and variances–covariances. *Psychometrika*, *83*, 425–442.
Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., & Shen, W. (2017). Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics*, *26*(1), 121–133.
Zhang, J., Lu, J., & Zhang, Z. (2025). Modeling missing response data in item response theory: Addressing missing not at random mechanism with monotone missing characteristics. *Journal of Educational Measurement*. https://doi.org/10.1111/jedm.12428

Cite this article: Wu, J., Ma, Z. and Chen, M.-H. (2026). Bayesian Modeling and Inference for Item Response Model with Nonignorable Missing Data. *Psychometrika*, 1280–1321. https://doi.org/10.1017/psy.2026.10121

https://doi.org/10.1017/psy.2026.10121 Published online by Cambridge University Press