BAYESIAN MODEL ASSESSMENT FOR JOINTLY MODELING MULTIDIMENSIONAL RESPONSE DATA WITH APPLICATION TO COMPUTERIZED TESTING

FANG LIU

NORTHEAST NORMAL UNIVERSITY

Xiaojing Wang, Roeland Hancock and Ming-Hui Chen

UNIVERSITY OF CONNECTICUT

March 21, 2022

Correspondence should be sent to

E-Mail: xiaojing.wang@uconn.edu
Phone: 860-486-3415
Fax: 860-486-4113
Website: https://xiaojing-wang.uconn.edu

BAYESIAN MODEL ASSESSMENT FOR JOINTLY MODELING MULTIDIMENSIONAL RESPONSE DATA WITH APPLICATION TO COMPUTERIZED TESTING

Abstract

Computerized assessment provides rich multidimensional data including trial-by-trial accuracy and response time (RT) measures. A key question in modeling this type of data is how to incorporate RT data, for example, in aid of ability estimation in item response theory (IRT) models. To address this, we propose a joint model consisting of a two-parameter IRT model for the dichotomous item response data, a log-normal model for the continuous RT data, and a normal model for corresponding pencil-and-paper scores. Then, we reformulate and reparameterize the model to capture the relationship between the model parameters, to facilitate the prior specification, and to make the Bayesian computation more efficient. Further, we propose several new model assessment criteria based on the decomposition of deviance information criterion (DIC) and the logarithm of the pseudo-marginal likelihood (LPML). The proposed criteria can quantify the improvement in the fit of one part of the multidimensional data given the other parts. Finally, we have conducted several simulation studies to examine the empirical performance of the proposed model assessment criteria, and have illustrated the application of these criteria using a real dataset from a computerized educational assessment program.

Key words: Computerized tests, DIC decomposition, IRT models, LPML decomposition, Pencil-and-paper tests, Response times.

1. Introduction

The current gold standard for educational and neuropsychological assessment is the use of pencil-and-paper assessments, individually administered by a trained assessor. Such tests are of immense value for the early identification and appropriate intervention of children at risk for learning disorders and long-term academic problems, which is a key part of a response-to-intervention treatment model (Gilbert et al., 2012). However, mass application of standard tests comes with considerable expenditures of both time and resources. The computerized assessment provides a potential solution, and provides rich multidimensional data including trial-by-trial accuracy and response time measures.

A case in point is the AppRISE data. In the AppRISE tablet assessment, participants begin each trial by pressing a start button to play an audio recording of a pseudoword (e.g., /feg/). An array of four printed response options (e.g., the target *feg*, and distractors *fep*, *fod*, and *fet*) is displayed and the participant selects the correct printed word to activate a machine and end the trial (see `https://haskinsglobal.org/apprise/` for more details). Both response accuracy and response time (RT) (from the option display) are collected for each item and can be used to model participant-level pseudoword reading skill. For comparison, the Phonemic Decoding Efficiency (PDE) subtest of the Test of Word Reading Efficiency-Second Edition (Torgesen et al., 2012) was also administered. In this pencil-and-paper test, participants read a list of regular pseudowords of increasing complexity as quickly as possible. The number of items read correctly within 45 seconds is converted to an age-normed standard score. Therefore, we have three types of data available in the study, which are the item response and RT from computerized assessment as well as standard score from paper-and-pencil PDE test.

To analyze the accuracy for item responses in the assessment, the item response theory (IRT) has become a popular method (Fox, 2010; van der Linden, 2017). With the increasing prevalence of computerized assessments, collecting RT has become much easier.

There is increasing attention in the field of IRT to including the RT data in modeling item responses. Several researchers have pointed out that the RT data is an important source of information to refine the inference on one's ability in a test (Luce, 1991; van der Linden, 2009). In fact, many current works have focused on the joint modeling of the item response and RT data. For instance, Entink et al. (2009) built up a multivariate multilevel regression for jointly modeling the item response and RT data, where they incorporated covariates for explaining the variation of speed and accuracy between individuals and groups of test takers. Loeys et al. (2011) proposed a Bayesian hierarchical framework for jointly modeling the item response and RT data, where they can estimate the correlation between speed and ability at the participant level. Wang et al. (2016) put forward a joint model of the item response and RT in the longitudinal setting. Molenaar and de Boeck (2018) built up a mixture model of the item response and RT data in order to describe the heterogeneity in the data. Recently, Lu et al. (2020) extended this mixture model with a higher-order structure of ability to detect participants' rapid guessing behavior.

However, the responses collected from AppRISE tablet assessment and the Test of Word Reading Efficiency Phonemic Decoding (PDE) paper-and-pencil test, including item responses, RTs, and PDE scores, are multidimensional data, thus the current state of the joint modeling for the item response and the RT data in the literature does not directly apply to this data. Hence, in this paper, we propose a new joint model that can capture the relationships among the item responses, RTs, and PDE scores in a coherent way. To be specific, we propose a trivariate normal distribution among the latent traits of each participant by adding a hierarchical layer upon the proposed joint model.

It is also important to ask whether certain parts of the multidimensional data can help, for example, the model fit of the item response data when we jointly analyze these different sources of information. From a practical perspective, it is important to understand how multiple assessments may aid in quantifying information gain in latent ability estimation so that assessments can be conducted efficiently in clinical and educational practice. In many papers relevant to the conjoint modeling of the RT and item response data, the deviance information criterion (DIC) (Spiegelhalter et al., 2002) is commonly used to evaluate the model fit (Johnson, 2003; Donkin et al., 2009; Rouder et al., 2015). For example, Entink et al. (2009) compared four different multilevel joint models using the DIC, and Loeys et al. (2011) applied the DIC to investigate joint modeling of the item response and RT data versus modeling them separately. But in the current literature, the DIC is seldom applied to assess the information gain in modeling the item response data by conjointly modeling the RT data. Although Wang et al. (2016) proposed a partial DIC idea to evaluate different RT models in contribution to the fit of the joint modeling, they assumed the latent ability in the joint model is given for the computation of the partial DIC. Since the latent ability is often unknown, their partial DIC lacks the power to fully quantify the information gain of incorporating RT in the joint model.

Thus, inspired by D. Zhang et al. (2017), we propose a new model assessment method, which enables us to fully quantify the information gain from each piece in the joint model. The proposed model assessment method will be applied to a dataset from a novel gamified tablet-based assessment (AppRISE) targeted towards assessing foundational pre-literacy, literacy, and cognitive skills in early education.

There are several contributions we have made in this paper. First, we reformulate and reparameterize the joint model to better describe multidimensional data from computerized testing, facilitate prior specification for the model parameters, and make posterior computations more efficient. Next, we propose a new Bayesian method for evaluating the contributions of the RT and PDE data to the fit of the item response data, the item response and PDE data to the fit of the RT data, as well as the RT and response data to the fit of the PDE data within the joint modeling framework. Specifically, we decompose the DIC and the logarithm of the pseudo-marginal likelihood (LPML) for the joint model into several additive components, which allows us to quantify the contributions of different parts of the data. The deviance functions in the DIC decomposition, the joint distributions, as well as the marginal distributions of the item responses, RTs, and PDEs involved in the LPML decomposition depend only on one-dimensional integrals after analytically integrating out latent speed variables. We then develop an efficient way using readily available Markov chain Monte Carlo (MCMC) samples from the joint posterior distribution to compute all quantities in the DIC and LPML decomposition. In addition, we carry out extensive simulation studies to examine the performance of the model in recovering the true parameters for the joint model and evaluate the empirical performance of the proposed model assessment criteria based on the DIC and LPML decompositions over different numbers of items and participants. Finally, we carry out residual analysis and item ranking for the multidimensional data and the results are given in the Supplementary Materials.

The rest of the article is organized as follows. In Section 2, we introduce the joint models of the item response, RT and PDE data. Section 3 presents the priors, the likelihood and the posterior distribution. Then, we develop the DIC decomposition and LPML decomposition in Section 4 and also discuss their computational details. In Section 5, we conduct two simulation studies. The AppRISE data is further analyzed to demonstrate the proposed methods in Section 6. In the Supplementary Materials, we propose some statistical tools that can be used in examining model adequacy and in ranking item difficulty while considering uncertainty. Finally, we summarize our major results and discuss some future directions of this work in Section 7.

2. The Proposed Method

2.1. Joint Model and Hierarchical Framework

In this section, we propose a joint model for the item response and the RT from the AppRISE tablet assessment, and the age-normed standard score of Phonemic Decoding Efficiency (PDE) from a paper-and-pencil test. Due to the test design of AppRISE, we use the two-parameter IRT model (van der Linden & Hambleton, 2013) for item accuracy. To account for the positive skew of the RT data, a log-normal distribution is specified to model the RTs (van der Linden & Guo, 2008; van der Linden, 2009; Fox & Marianti, 2016; Man et al., 2019). A normal model is assumed for the standardized PDE score. Finally, we assume a trivariate normal distribution for the latent ability parameter, the latent speed parameter, and the error term of the PDE score to capture the association between the AppRISE tablet assessment and the paper-based PDE score as well as to borrow the strength of information among different types of data.

Let $y_{ij(i)}$ denote the binary response for the ith participant taking the $j(i)$th item, having a value of "0" or "1", $t_{ij(i)}$ be the response time in seconds for the ith participant answering the $j(i)$th, and PDE_i indicate the paper-based PDE score for the ith participant for $i = 1, \cdots, N$ and $j(i) \in \{1, \cdots, J\}$ with J as the maximum number of items administrated in the test. Notice that the subscript $ij(i)$ is a nested structure, which allows each participant to take a different number of items in a test. This is very common in computerized test to decrease administration time in educational settings. To simplify the notation in the subscript, we will suspend the (i) in the subscript throughout the paper without causing any further confusion.

Then, the joint model is given by

$$p_{ij} = \text{P}(y_{ij} = 1 \mid a_j, b_j, \theta_i^*) = \text{F}\left[a_j(\theta_i^* - b_j)\right] = \frac{\exp\{a_j(\theta_i^* - b_j)\}}{1 + \exp\{a_j(\theta_i^* - b_j)\}}, \tag{2.1}$$

$$\log t_{ij} = \lambda_j - \tau_i^* + \epsilon_{ij}, \tag{2.2}$$

$$\text{PDE}_i = \beta_0 + \zeta_i^*, \tag{2.3}$$

and

$$(\theta_i^*, \tau_i^*, \zeta_i^*)' \overset{i.i.d.}{\sim} \mathcal{N}_3\left((0,0,0)', \mathbf{\Sigma}\right) \text{ with } \mathbf{\Sigma} = \begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_\tau^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}. \tag{2.4}$$

In equation 2.1, p_{ij} is the probability that the ith participant answers the jth item correctly, $\text{F}(\cdot)$ is a continuous cumulative distribution function (cdf), θ_i^* is the ith participant's latent ability, and a_j and b_j are item discrimination and item difficulty, respectively, of the jth item. We choose $\mathrm{F}(\cdot)$ to be a logistic link function in equation 2.1 for our study. In equation 2.2, τ_i^* denotes the speed (the participant uses to complete items during a test) of the ith participant, λ_j indicates the time intensity required (van der Linden, 2009) for the jth item, ϵ_{ij} is an error term following a normal distribution with mean centered at zero and variance σ_j^2, and $1/\sigma_j^2$ quantifies the dispersion of the lognormal distribution, implying the discrimination power of the jth item. In equation 2.3, β_0 can be regarded as the average PDE score across participants. Finally, in equation 2.4, we assume $(\theta_i^*, \tau_i^*, \zeta_i^*)'$ follows a trivariate normal distribution, the variance of θ_i^* is assumed to be 1 to ensure identifiability, σ_τ^2 is the variance of τ_i^*, σ_{12}/σ_τ captures the correlation between θ_i^* and τ_i^*, σ_{33} is the variance of ζ_i^*, σ_{13} is the covariance between θ_i^* and ζ_i^*, and σ_{23} is the covariance between τ_i^* and ζ_i^*.

2.2. Reformulation and Reparameterization

Since Σ in equation 2.4 has to be a positive definite covariance matrix, the specification of the prior distributions for σ_{12}, σ_τ^2, σ_{13}, σ_{23}, and σ_{33} is not straightforward. In addition, sampling these parameters from the posterior distribution is challenging. Therefore, we employ a reformulation and reparameterization to facilitate specification of the prior distributions and to allow a convenient and more efficient implementation of Bayesian computation.

First, let us consider the following reformulation

$$(\theta_i^*, \tau_i^*, \zeta_i^*)' = \mathbf{\Gamma}\,(\theta_i, \tau_i, \zeta_i)' \text{ with } \mathbf{\Gamma} = \begin{pmatrix} 1 & 0 & 0 \\ \sigma_\tau \sin\varphi & \sigma_\tau \cos\varphi & 0 \\ \beta_1 & \beta_2 & 1 \end{pmatrix},$$
$$(\theta_i, \tau_i, \zeta_i)' \overset{i.i.d.}{\sim} \mathcal{N}_3\left((0,0,0)', \mathrm{diag}(1,1,\sigma^2_{\text{PDE}})\right). \tag{2.5}$$

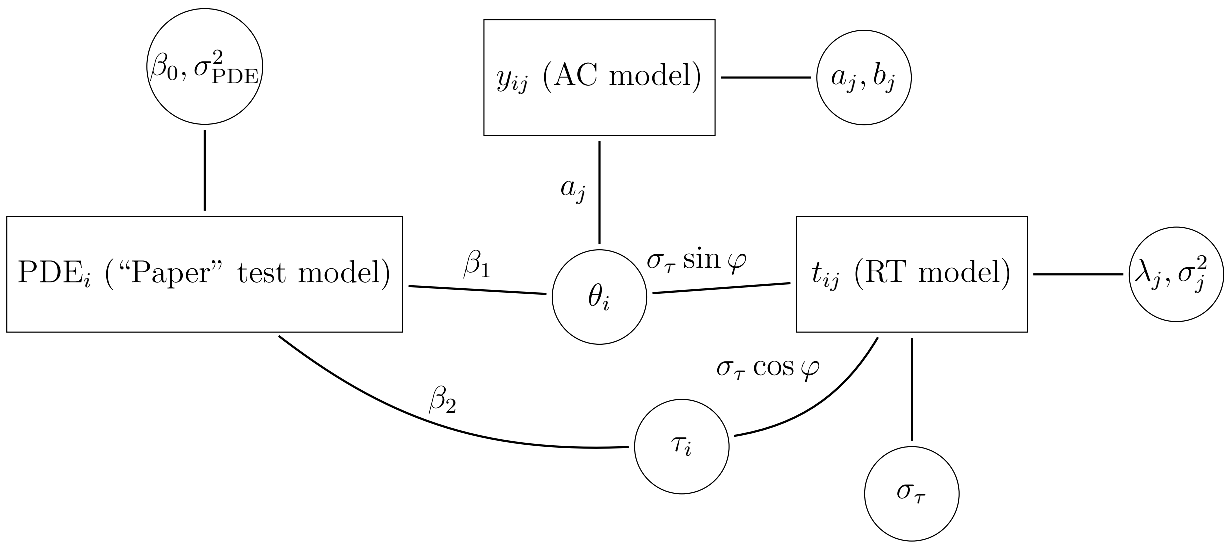

Figure 2.1: Relationships of the three model components in the joint model.

Then, we take the reparameterization

$$\begin{aligned}\sigma_{12} &= \sigma_\tau \sin\varphi,\\ \sigma_{13} &= \beta_1,\\ \sigma_{23} &= \sigma_\tau(\beta_1 \sin\varphi + \beta_2 \cos\varphi),\\ \sigma_{33} &= \beta_1^2 + \beta_2^2 + \sigma^2_{\text{PDE}},\\ \sigma_\tau &= \exp(w).\end{aligned} \tag{2.6}$$

With reformulation equation 2.5 and reparameterization equation 2.6, the joint distribution of $(\theta_i^*, \tau_i^*, \zeta_i^*)'$ follows a trivariate normal distribution given in equation 2.4. We summarize the three components of our joint model after the reformulation and reparameterization as

$$\textit{AC model}: \quad p_{ij} = P(y_{ij} = 1 \mid a_j, b_j, \theta_i) = \frac{\exp\{a_j(\theta_i - b_j)\}}{1 + \exp\{a_j(\theta_i - b_j)\}}, \tag{2.7}$$

$$\textit{RT model}: \quad \log t_{ij} = \lambda_j - \sigma_\tau(\theta_i \sin\varphi + \tau_i \cos\varphi) + \epsilon_{ij}, \tag{2.8}$$

$$\textit{“Paper” test model}: \quad \text{PDE}_i = \beta_0 + \beta_1\theta_i + \beta_2\tau_i + \zeta_i. \tag{2.9}$$

Here, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_j^2)$ and $\zeta_i \sim \mathcal{N}(0, \sigma^2_{\text{PDE}})$, which are independent of θ_i and τ_i. Figure 2.1

presents the relationships among the model components in equation 2.7 to equation 2.9, where the rectangle block indicates the model and its corresponding observations; the circle indicates the unknown parameters; and the line between the circle and the rectangle implies how the model and parameters relate to each other. It is apparent that the reparameterization makes the interpretation for the relations of these model components much simpler, where θ_i explicitly ties up all three model components, while τ_i obviously connects the RT model and the "paper" test model. After the reparameterization, in equation 2.6, $-\infty < \beta_1 < \infty$, $-\infty < \beta_2 < \infty$, $\sigma^2_{\text{PDE}} > 0$, $-\infty < w < \infty$, and $-\pi/2 < \varphi < \pi/2$. Unlike Σ in equation 2.4, there are no other constraints on the parameters β_1, β_2, σ^2_{PDE}, and φ. In addition, $\beta_1 = 0$ implies $\sigma_{13} = 0$, indicating independence between the item response and PDE, while the response time and PDE are independent if $\beta_1 = \beta_2 = 0$ or $\varphi = \beta_2 = 0$.

3. The Priors, Likelihood and Posterior Distribution

Due to the complexity of the joint model, we resort to Bayesian computation for the inference of unknown parameters. First, we need to assume priors for all unknown parameters. For the item discrimination parameter a_j of the jth item, we assume it follows a normal distribution truncated by zero at the left with its mean centered at 0 and variance being 1, i.e., $a_j \sim \mathcal{N}_+(0,1)$ (c.f., Karadavut (2019)); for the item difficulty parameter b_j, we give a hierarchical normal prior $\mathcal{N}(\mu_b, \sigma_b^2)$, where $\mu_b \sim \mathcal{N}(0, 10\sigma_b^2)$ and σ_b^2 follows an inverse gamma distribution, i.e., $IG(0.1, 0.1)$. As discussed in Section 2.2, there are no constraints on the four reparameterized parameters β_1, β_2, σ^2_{PDE}, w, and φ. Thus, it is reasonable to assume independent priors for these parameters. Specifically, we assume $\beta_1 \sim \mathcal{N}(0, 10^2)$, $\beta_2 \sim \mathcal{N}(0, 10^2)$, $\sigma^2_{\text{PDE}} \sim IG(0.1, 0.1)$, $w \sim N(0, 10^2)$, and $\varphi \sim U(-\pi/2, \pi/2)$. For the intercept in the "paper" test model, we further assume $\beta_0 \sim \mathcal{N}(0, 10^2)$. Similarly, we specify an inverse gamma prior, $\sigma_j^2 \sim IG(0.1, 0.1)$, for the variance σ_j^2 and a normal prior, $\lambda_j \sim \mathcal{N}(0, 10^2)$, for the time intensity parameter λ_j in the RT model.

Next, we establish the joint posterior distribution of all model parameters in the proposed joint model. Let $\boldsymbol{a} = (a_1, \cdots, a_J)'$, $\boldsymbol{b} = (b_1, \cdots, b_J)'$, $\boldsymbol{\lambda}_J = (\lambda_1, \cdots, \lambda_J)'$, $\boldsymbol{\gamma}_1 = (\boldsymbol{a}', \boldsymbol{b}')'$, $\boldsymbol{\gamma}_2 = (\beta_0, \beta_1, \beta_2, \sigma^2_{\text{PDE}}, \boldsymbol{\lambda}'_J, \sigma_\tau, \varphi, \sigma^2_1, \cdots, \sigma^2_J)'$. Write $\boldsymbol{\gamma}_2 = (\boldsymbol{\gamma}'_{21}, \boldsymbol{\gamma}'_{22})'$ with $\boldsymbol{\gamma}_{21} = (\beta_0, \beta_1, \beta_2, \sigma^2_{\text{PDE}})'$ being the vector of parameters in the "paper" test model and $\boldsymbol{\gamma}_{22} = (\boldsymbol{\lambda}'_J, \varphi, \sigma_\tau, \sigma^2_1, \cdots, \sigma^2_J)'$ being the vector of parameters in the RT model. Then, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_2)' = (\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_{21}, \boldsymbol{\gamma}'_{22})'$ is a vector of all unknown parameters except unknown participant parameters in the model. Further, denote $\boldsymbol{t}_i = (\log t_{i1}, \cdots, \log t_{iJ_i})'$ and $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iJ_i})'$ are the vectors for the RTs and item responses observed for each participant, respectively, where J_i denotes the total number of the items that have been administrated for the ith participant with $J_i \in \{1, \cdots, J\}$. Thus, given $\boldsymbol{\gamma}_1$ and θ_i, the conditional likelihood of $\boldsymbol{y}_i$ in the AC model is

$$\mathcal{L}(\boldsymbol{\gamma}_1, \theta_i \mid \boldsymbol{y}_i) = f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i) = \prod_{j=1}^{J_i} \frac{\exp\{y_{ij} a_j(\theta_i - b_j)\}}{1 + \exp\{a_j(\theta_i - b_j)\}}, \tag{3.1}$$

while given the parameters $\boldsymbol{\gamma}_{22}$, θ_i and τ_i, the conditional likelihood of $\boldsymbol{t}_i$ is

$$\mathcal{L}(\boldsymbol{\gamma}_{22}, \theta_i, \tau_i \mid \boldsymbol{t}_i) = f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22}, \theta_i, \tau_i) = \prod_{j=1}^{J_i} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{[\log t_{ij} - \lambda_j + \sigma_\tau(\theta_i \sin\varphi + \tau_i \cos\varphi)]^2}{2\sigma^2_j}\}.$$

Assuming that given latent variables θ_i and τ_i for the ith participant, the observations of item responses, RTs and PDE scores are conditionally independent. Then, the joint probability density function (pdf) of $\mathcal{D}_{i,obs} = (\boldsymbol{y}'_i, \boldsymbol{t}'_i, \text{PDE}_i)'$ is written as

$$f(\mathcal{D}_{i,obs} \mid \boldsymbol{\gamma}, \theta_i, \tau_i) = f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i) f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22}, \theta_i, \tau_i) f(\text{PDE}_i \mid \boldsymbol{\gamma}_{21}, \theta_i, \tau_i), \tag{3.2}$$

where $f(\text{PDE}_i \mid \boldsymbol{\gamma}_{21}, \theta_i, \tau_i) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(\text{PDE}_i - \beta_0 - \beta_1\theta_i - \beta_2\tau_i)^2}{2\sigma^2_{\text{PDE}}}\}$. Next, by integrating out latent variables θ_i and τ_i, we obtain the marginal likelihood of $\boldsymbol{\gamma}$ as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}) &= \prod_{i=1}^{N} f(\boldsymbol{y}_i, \boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}) \\ &= \prod_{i=1}^{N} \int\int f(\boldsymbol{y}_i, \boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}, \theta_i, \tau_i)\phi(\theta_i)\phi(\tau_i)d\theta_i d\tau_i, \qquad (3.3)\\ &= \prod_{i=1}^{N} \int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i)\left[\int \phi(\tau_i) f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22}, \theta_i, \tau_i) f(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{21}, \theta_i, \tau_i)d\tau_i\right] d\theta_i, \qquad (3.4)\end{aligned}$$

where $\mathcal{D}_{obs} = (\mathcal{D}'_{1,obs}, \cdots, \mathcal{D}'_{N,obs})'$ and $\phi(\cdot)$ denotes the standard normal pdf, which represents the prior assigned to θ_i and τ_i. In practice, the high dimensional numerical integration is often expensive. However, in our case, we can reduce the two-dimensional integration in equation 3.3 to one-dimensional integration as shown in equation 3.4. Based on the normality of τ_i, $\boldsymbol{t}_i$ and PDE_i, it turns out that the second integrand of equation 3.4 is also a normal pdf and makes the integration of τ_i analytically tractable. After integrating out the τ_i in closed form, the remaining component only depends on θ_i. Thus, it is easy to numerically compute the one-dimensional integration of θ_i in equation 3.4.

Once the marginal likelihood of $\boldsymbol{\gamma}$ is computed, the joint posterior of $\boldsymbol{\gamma}$ is given by

$$\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}) = \frac{\mathcal{L}(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})\pi(\boldsymbol{\gamma})}{c(\mathcal{D}_{obs})}, \qquad (3.5)$$

where $\pi(\boldsymbol{\gamma}) = \pi(\beta_0)\pi(\beta_1)\pi(\beta_2)\pi(\sigma^2_{\mathrm{PDE}})\pi(\sigma_\tau)\pi(\varphi)\left[\prod_{j=1}^{J} \pi(a_j)\pi(b_j \mid \mu_b, \sigma_b^2)\pi(\lambda_j)\pi(\sigma_j^2)\right] \times \pi(\mu_b)\pi(\sigma_b^2)$ is the joint prior of the unknown parameter vector $\boldsymbol{\gamma}$ and the prior of each unknown parameter in this product is specified at the beginning of this section. Further, in equation 3.5, $c(\mathcal{D}_{obs}) = \int \prod_{i=1}^{N} f(\boldsymbol{y}_i, \boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma})\pi(\boldsymbol{\gamma})d\boldsymbol{\gamma}$ is the normalized constant. Directly sampling from equation 3.5 is somewhat difficult, thus, we propose to use the augmented joint posterior density function of $\boldsymbol{\gamma}$, $\boldsymbol{\theta}^R$ and $\boldsymbol{\tau}^R$ as below in our MCMC computation,

$$\pi(\boldsymbol{\gamma}, \boldsymbol{\theta}^R, \boldsymbol{\tau}^R \mid \mathcal{D}_{obs}) = \frac{\prod_{i=1}^{N} f(\mathcal{D}_{i,obs} \mid \boldsymbol{\gamma}, \theta_i, \tau_i)\pi(\theta_i)\pi(\tau_i)\pi(\boldsymbol{\gamma})}{c(\mathcal{D}_{obs})},$$

where $f(\mathcal{D}_{i,obs} \mid \boldsymbol{\gamma}, \theta_i, \tau_i)$ is defined in equation 3.2, $\boldsymbol{\theta}^R = (\theta_1, \cdots, \theta_N)'$ and $\boldsymbol{\tau}^R = (\tau_1, \cdots, \tau_N)'$. Then, we can show that $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$ is the marginal posterior distribution of $\pi(\boldsymbol{\gamma}, \boldsymbol{\theta}^R, \boldsymbol{\tau}^R \mid \mathcal{D}_{obs})$ by integrating out $\boldsymbol{\theta}^R$ and $\boldsymbol{\tau}^R$.

The reformulation and reparametrization in Section 2.2 leads to a convenient and efficient implementation of MCMC sampling from the joint posterior distribution in equation 3.5 using an R package called *nimble* (de Valpine et al., 2017, 2020). In *nimble*, we use the calling command `getSampler()` to sample $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{\theta}$, φ, and σ_τ via the Metropolis-Hastings adaptive random-walk sampler and draw $\boldsymbol{\tau}$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\lambda}_J$, σ_b^2, σ^2_{PDE}, β_0, β_1, β_2, and μ_b via the Gibbs sampler. Once MCMC samples are generated from *nimble*, we compute the posterior means, posterior standard deviations, and 95% highest posterior density (HPD) intervals for all of the model parameters using an R package *boa*. Then, using the MCMC samples from *nimble*, we develop program codes written in Fortran 95 and Matlab to compute all of the proposed model assessment criteria in Section 4. The detailed codes for *nimble*, Fortran, and Matlab are provided in Section S.6 of the Supplementary Materials.

4. Model Assessment Criteria

In this section, we propose decomposition of two commonly used model selection criteria to assess the contribution of any two parts of the multidimensional data in the analysis of the remaining part of the data. The model assessment criteria we focus on are the deviance information criterion (DIC) (Spiegelhalter et al., 2002) and a conditional predictive ordinate (CPO) (Geisser & Eddy, 1979; Gelfand et al., 1992; Gelfand & Dey, 1994) related criterion called the logarithm of the pseudo-marginal likelihood (LPML) criterion (Ibrahim et al., 2001).

4.1. Deviance Information Criterion

The use of DIC for model comparison is abundant in the psychometric literature. Examples of its use in joint modeling of RT and item response data include Johnson (2003), Entink et al. (2009), Rouder et al. (2015), Bolsinova et al. (2017), and Lu et al. (2020). Typically, when we perform Bayesian inference on latent variable models, we directly sample the latent variables along with the model parameters. For convenience, the conditional DIC is often constructed via the conditional likelihood in which the latent variables are treated as "model parameters".

However, several studies show that conditional DICs usually have much larger Monte Carlo errors compared to the DICs based on the marginal likelihoods by integrating out latent variables (Celeux et al., 2006; Chan & Grant, 2016; Merkle et al., 2019; X. Zhang et al., 2019). For example, X. Zhang et al. (2019) found that the conditional DIC generally selects a model that is more complex than the true model. Li et al. (2020) discussed the problems of the conditional DICs in terms of the dimension of the parameter space, frequentist justification, and asymptotic properties. They also proposed a new version of DIC, called DIC_L, where they studied the large sample properties of DIC_L, and introduced the expectation–maximization (EM) algorithm, Kalman and particle filtering algorithms to compute this new DIC for latent variable models. DIC_L is potentially useful for evaluating complex models such as the ones considered in this paper. Moreover, in the paper of Merkle et al. (2019), they pointed out the model assessment criteria based on the information of marginal likelihoods could be used to evaluate the predictive ability of a model when it was applied to new clusters (e.g., countries, schools, or districts), which is a desirable feature in many psychometric contexts, as we often wish to distinguish general properties of items that are not specific to what we observed.

Hence, we define the DIC of the proposed joint model as

$$\text{DIC} = \text{Dev}(\overline{\boldsymbol{\gamma}}) + 2p_D, \tag{4.1}$$

where $\overline{\boldsymbol{\gamma}}$ is the posterior mean of $\boldsymbol{\gamma}$, and $p_D = \mathrm{E}_{\pi(\boldsymbol{\gamma}|\mathcal{D}_{obs})}\left[\mathrm{Dev}(\boldsymbol{\gamma}) \mid \mathcal{D}_{obs}\right] - \mathrm{Dev}(\overline{\boldsymbol{\gamma}})$ is the effective number of model parameters. Here, for p_D, we take the expectation regarding to the posterior distribution of $\boldsymbol{\gamma}$ and in practice, to compute $\mathrm{E}_{\pi(\boldsymbol{\gamma}|\mathcal{D}_{obs})}\left[\mathrm{Dev}(\boldsymbol{\gamma}) \mid \mathcal{D}_{obs}\right]$, we often take the posterior mean of $\mathrm{Dev}(\boldsymbol{\gamma})$, where $\mathrm{Dev}(\boldsymbol{\gamma}) = -2\log\mathcal{L}(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$ is the deviance function with the marginal likelihood $\mathcal{L}(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$ defined in equation 3.3. Thus, for DIC in equation 4.1, we integrate out all latent parameters related to a participant, such as θ_i and τ_i, in the likelihood function. Since the deviance function $\mathrm{Dev}(\boldsymbol{\gamma})$ only depends on $\mathcal{L}(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$, which, as discussed in Section 3, needs to compute an one-dimensional integral where we compute directly using the global adaptive quadrature method (Visual Numerics, 2003).

4.1.1. DIC Decomposition

First, let us focus on the part of only modeling the RT and PDE data in the joint model. With some algebra, given the parameters $\boldsymbol{\gamma}_2$ (only involved in the RT and PDE model), $(\theta_i, \boldsymbol{t}_i', \mathrm{PDE}_i)'$ follows a (J_i+2)-dimensional multivariate normal distribution, i.e.,

$$[(\theta_i, \boldsymbol{t}_i', \mathrm{PDE}_i)' \mid \boldsymbol{\gamma}_2] \sim \mathcal{N}_{J_i+2}\left(\left(0, \boldsymbol{\lambda}_{J_i}', \beta_0\right)', \boldsymbol{\Sigma}_i\right) \tag{4.2}$$

$$\text{with} \quad \boldsymbol{\Sigma}_i = \begin{pmatrix} 1 & -\sigma_\tau \sin\varphi \mathbf{1}_{J_i}' & \beta_1 \\ -\sigma_\tau \sin\varphi \mathbf{1}_{J_i} & \sigma_\tau^2 \mathbf{1}_{J_i}\mathbf{1}_{J_i}' + \bigoplus_{j=1}^{J_i} \sigma_j^2 & -(\beta_1 \sin\varphi + \beta_2 \cos\varphi)\sigma_\tau \mathbf{1}_{J_i} \\ \beta_1 & -(\beta_1 \sin\varphi + \beta_2\cos\varphi)\sigma_\tau \mathbf{1}_{J_i}' & \sigma_{\mathrm{PDE}}^2 + \beta_1^2 + \beta_2^2 \end{pmatrix},$$

where $\boldsymbol{\lambda}_{J_i} = (\lambda_1, \cdots, \lambda_{J_i})'$; $\mathbf{1}_{J_i}$ is a J_i-dimensional column vector with all elements being 1 and $\mathbf{1}_{J_i}'$ is its transpose; $\bigoplus_{j=1}^{J_i}$ is the direct sum operator and $\bigoplus_{j=1}^{J_i} \sigma_j^2$ is a $J_i \times J_i$ diagonal matrix with the (j,j)th entry equal to σ_j^2. By the marginal property of the multivariate normal distribution and equation 4.2, the pdf $f(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2)$ is also a multivariate normal pdf.

Then, the deviance function $\mathrm{Dev}_{[\mathrm{RT,PDE}]}(\boldsymbol{\gamma}_2) = -2\sum_{i=1}^{N} \log f(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2)$ is just defined for jointly modeling the RT and PDE data; further, the DIC of the RT and PDE parts is

$$\mathrm{DIC}_{[\mathrm{RT,PDE}]} = \mathrm{Dev}_{[\mathrm{RT,PDE}]}(\overline{\boldsymbol{\gamma}}_2) + 2p_{D[\mathrm{RT,PDE}]},$$

where the corresponding effective number of model parameters $p_{D[\mathrm{RT,PDE}]}$ equals to $\mathrm{E}_{\pi(\boldsymbol{\gamma}_2|\mathcal{D}_{obs})}\left[\mathrm{Dev}_{[\mathrm{RT,PDE}]}(\boldsymbol{\gamma}_2) \mid \mathcal{D}_{obs}\right] - \mathrm{Dev}_{[\mathrm{RT,PDE}]}(\overline{\boldsymbol{\gamma}}_2)$, $\overline{\boldsymbol{\gamma}}_2$ is the posterior mean of $\boldsymbol{\gamma}_2$ and the expectation $\mathrm{E}_{\pi(\boldsymbol{\gamma}_2|\mathcal{D}_{obs})}$ is now taken with respect to the posterior distribution of $\boldsymbol{\gamma}_2$, as only the parameter $\boldsymbol{\gamma}_2$ is involved in the model of the RT and PDE parts. Similarly, the deviance functions for jointly modeling the RT and item response data as well as the item response and PDE data are defined as $\mathrm{Dev}_{[\mathrm{AC,RT}]}(\boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1) = -2\sum_{i=1}^{N} \log f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1)$ and $\mathrm{Dev}_{[\mathrm{AC,PDE}]}(\boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1) = -2\sum_{i=1}^{N} \log f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1)$, respectively, where the likelihood functions $f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1) = \int\int f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22}, \theta_i, \tau_i) f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i)\phi(\tau_i)d\theta_i d\tau_i$ and $f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1) = \int\int f(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{21}, \theta_i, \tau_i) f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i)\phi(\tau_i)d\theta_i d\tau_i$, respectively. Notice that the two-dimensional integrals in term of θ_i and τ_i in these likelihoods can be reduced to one-dimensional integrals as shown in equation 4.3 and equation 4.4.

Proposition 1. The joint densities of $f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1)$ and $f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1)$ can be expressed as the functions of one-dimensional integrals with respect to θ_i given by

$$f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1) = \int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i) f_N(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_2, \theta_i)d\theta_i, \tag{4.3}$$

$$f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1) = \int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i) f_N(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{21}, \theta_i)d\theta_i, \tag{4.4}$$

where $f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)$ is given in equation 3.1, $f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22}, \theta_i)$ is the pdf of a multivariate normal distribution $\mathcal{N}_{J_i}(\boldsymbol{\lambda}_{J_i} - \sigma_\tau \theta_i \sin\varphi \mathbf{1}_{J_i}, \boldsymbol{\Sigma}_{t_i})$ with $\boldsymbol{\Sigma}_{t_i} = \sigma_\tau^2 \mathbf{1}_{J_i}\mathbf{1}'_{J_i} + \bigoplus_{j=1}^{J_i} \sigma_j^2 - \sigma_\tau^2 \sin^2\varphi \mathbf{1}_{J_i}\mathbf{1}'_{J_i}$, while $f_N(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{21}, \theta_i)$ is the pdf of a normal distribution $\mathcal{N}(\beta_0 + \beta_1\theta_i, \beta_2^2 + \sigma_{PDE}^2)$.

Then, $\mathrm{DIC}_{[\mathrm{AC,PDE}]} = \mathrm{Dev}_{[\mathrm{AC,PDE}]}(\overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{21}) + 2p_{D[\mathrm{AC,PDE}]}$ is the DIC of the AC and PDE parts, where the corresponding effective number of model parameters $p_{D[\mathrm{AC,PDE}]}$ equals to $\mathrm{E}_{\pi(\boldsymbol{\gamma}_{21},\boldsymbol{\gamma}_1|\mathcal{D}_{\mathrm{obs}})}[\mathrm{Dev}_{[\mathrm{AC,PDE}]}(\boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1) \mid \mathcal{D}_{\mathrm{obs}}] - \mathrm{Dev}_{[\mathrm{AC,PDE}]}(\overline{\boldsymbol{\gamma}}_{21}, \overline{\boldsymbol{\gamma}}_1)$, $\overline{\boldsymbol{\gamma}}_{21}$ and $\overline{\boldsymbol{\gamma}}_1$ are the posterior means, and the expectation $\mathrm{E}_{\pi(\boldsymbol{\gamma}_{21},\boldsymbol{\gamma}_1|\mathcal{D}_{\text{obs}})}$ is taken with respect to the posterior distribution of $\boldsymbol{\gamma}_{21}$ and $\boldsymbol{\gamma}_1$. Similarly, the DIC of AC and PDE parts is defined as $\mathrm{DIC}_{[\mathrm{AC,RT}]} = \mathrm{Dev}_{[\mathrm{AC,RT}]}(\overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{22}) + 2p_{D[\mathrm{AC,RT}]}$, where the corresponding effective number of model parameter $p_{D[\mathrm{AC,RT}]} = \mathrm{E}_{\pi(\boldsymbol{\gamma}_{22},\boldsymbol{\gamma}_1|\mathcal{D}_{\text{obs}})}[\mathrm{Dev}_{[\mathrm{AC,RT}]}(\boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1) \mid \mathcal{D}_{\text{obs}}] - \mathrm{Dev}_{[\mathrm{AC,RT}]}(\overline{\boldsymbol{\gamma}}_{22}, \overline{\boldsymbol{\gamma}}_1)$, $\overline{\boldsymbol{\gamma}}_{22}$ is the posterior mean, and the expectation $\mathrm{E}_{\pi(\boldsymbol{\gamma}_{22},\boldsymbol{\gamma}_1|\mathcal{D}_{\text{obs}})}$ is taken with respect to the posterior distribution of $\boldsymbol{\gamma}_{22}$ and $\boldsymbol{\gamma}_1$.

Since $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}) = \pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_{22} \mid \mathcal{D}_{obs})$, the Markov chain Monte Carlo (MCMC) samples of $\boldsymbol{\gamma}_2$, $(\boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1)$ and $(\boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1)$ from $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$ will be the same as drawn directly from their corresponding marginal distributions $\pi(\boldsymbol{\gamma}_2 \mid \mathcal{D}_{obs})$, $\pi(\boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1 \mid \mathcal{D}_{obs})$, and $\pi(\boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1 \mid \mathcal{D}_{obs})$. Hence, there are no additional MCMC draws needed for estimating $\mathrm{E}_{\pi(\boldsymbol{\gamma}_2|\mathcal{D}_{obs})}\left[\mathrm{Dev}_{[\mathrm{RT,PDE}]}(\boldsymbol{\gamma}_2) \mid \mathcal{D}_{obs}\right]$, $\mathrm{E}_{\pi(\boldsymbol{\gamma}_{22},\boldsymbol{\gamma}_1|\mathcal{D}_{\text{obs}})}[\mathrm{Dev}_{[\mathrm{AC,RT}]}(\boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1) \mid \mathcal{D}_{\text{obs}}]$, and $\mathrm{E}_{\pi(\boldsymbol{\gamma}_{21},\boldsymbol{\gamma}_1|\mathcal{D}_{\text{obs}})}[\mathrm{Dev}_{[\mathrm{AC,PDE}]}(\boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1) \mid \mathcal{D}_{\text{obs}}]$.

Next, based on the conditional property of the multivariate normal distribution, θ_i given $\boldsymbol{t}_i, \mathrm{PDE}_i$ and $\boldsymbol{\gamma}_2$ follows a normal distribution, i.e.,

$$[\theta_i \mid \boldsymbol{t_i}, \mathrm{PDE}_i, \boldsymbol{\gamma}_2] \ \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}), \tag{4.5}$$

where $\bar{\boldsymbol{\mu}} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left[(\boldsymbol{t}_i', \mathrm{PDE}_i)' - (\boldsymbol{\lambda}_{J_i}', \beta_0)'\right]$, $\bar{\boldsymbol{\Sigma}} = 1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, $\boldsymbol{\Sigma}_{12} = (-\sigma_\tau \sin\varphi \mathbf{1}_{J_i}', \beta_1)$, and $\boldsymbol{\Sigma}_{21}'$ is a transpose of $\boldsymbol{\Sigma}_{12}$, and

$$\boldsymbol{\Sigma}_{22} = \begin{pmatrix} \sigma_\tau^2 \mathbf{1}_{J_i}\mathbf{1}_{J_i}' + \bigoplus_{j=1}^{J_i} \sigma_j^2 & -(\beta_1 \sin\varphi + \beta_2 \cos\varphi)\sigma_\tau \mathbf{1}_{J_i} \\ -(\beta_1 \sin\varphi + \beta_2 \cos\varphi)\sigma_\tau \mathbf{1}_{J_i}' & \sigma_{\mathrm{PDE}}^2 + \beta_1^2 + \beta_2^2 \end{pmatrix}.$$

Equation 4.5 can be viewed as the prior information of θ_i obtained from the RT and PDE data given the parameter $\boldsymbol{\gamma}_2$. From equation 4.2, we see that for modeling the RT and PDE data to be independent of θ_i, it requires that both φ and β_1 are zero simultaneously, which can be further verified by equation 4.5. This can be better explained by Figure 2.1, when $\varphi = 0$, although there is no direct linkage between θ_i and the RT model, the influence of modeling RT for the estimation of θ_i can be transmitted through τ_i.

By putting the prior information of θ_i in equation 4.5 into the AC model, we can derive a deviance function of item responses data by given the information from the RT and PDE data as $\mathrm{Dev}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]}(\boldsymbol{\gamma}) = -2\sum_{i=1}^{N}\log f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ for $\boldsymbol{\gamma}$, where

$$f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \int f(\boldsymbol{y}_i \mid \theta_i, \boldsymbol{\gamma}_1) f(\theta_i \mid \boldsymbol{t}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_2) d\theta_i \tag{4.6}$$

with the pdf $f(\theta_i \mid \boldsymbol{t}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_2)$ defined in equation 4.5; further, we have the DIC of the AC given the RT and PDE as

$$\mathrm{DIC}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} = \mathrm{Dev}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]}(\overline{\boldsymbol{\gamma}}) + 2p_{D[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]}, \tag{4.7}$$

where $p_{D[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} = \mathrm{E}_{\pi(\boldsymbol{\gamma}|\mathcal{D}_{obs})}\left[\mathrm{Dev}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]}(\boldsymbol{\gamma}) \mid \mathcal{D}_{obs}\right] - \mathrm{Dev}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]}(\overline{\boldsymbol{\gamma}})$ is its corresponding effective number of parameters. Likewise, the deviance function of the RT data by given the information from the item responses and PDE data, as well as the deviance function of the PDE data by given the information from the item responses and the RT data are defined as $\mathrm{Dev}_{[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]}(\boldsymbol{\gamma}) = -2\sum_{i=1}^{N}\log f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \mathrm{PDE}_i, \boldsymbol{\gamma})$ and $\mathrm{Dev}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]}(\boldsymbol{\gamma}) = -2\sum_{i=1}^{N}\log f(\mathrm{PDE}_i \mid \boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\gamma})$, respectively, where the likelihood functions $f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \int\int f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22}, \theta_i, \tau_i) f(\theta_i, \tau_i \mid \boldsymbol{y}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21}) d\theta_i d\tau_i$ and $f(\mathrm{PDE}_i \mid \boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \int\int f(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{21}, \theta_i, \tau_i) f(\theta_i, \tau_i \mid \boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}) d\theta_i d\tau_i$. After some algebra calculations, these two-dimensional integrals in the likelihoods can be reduced to one-dimensional integrals. We formally state these results in Proposition 2.

Proposition 2. The conditional densities $f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ *and* $f(\mathrm{PDE}_i \mid \boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ *can be written as the ratios of two one-dimensional integrals, i.e,*

$$f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{\int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i) f_N(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2, \theta_i) d\theta_i}{f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1)}, \tag{4.8}$$

$$f(\mathrm{PDE}_i \mid \boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{\int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i) f_N(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2, \theta_i) d\theta_i}{f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1)}, \tag{4.9}$$

where $f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1)$ *and* $f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1)$ *are given in Proposition 1, while* $f_N(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2, \theta_i)$ *is the pdf of a multivariate normal distribution* $\mathcal{N}(\boldsymbol{\mu}_{tpde}, \boldsymbol{\Sigma}_{tpde})$ with $\boldsymbol{\mu}_{tpde} = (\boldsymbol{\lambda}'_{J_i}, \beta_0)' + \theta_i\boldsymbol{\Sigma}_{21}$, and $\boldsymbol{\Sigma}_{tpde} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{12}$.

Similarly, we can define the DIC for the RT given the AC and PDE as

$$\mathrm{DIC}_{[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]} = \mathrm{Dev}_{[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]}(\overline{\boldsymbol{\gamma}}) + 2p_{D[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]}, \tag{4.10}$$

with $p_{D[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]} = \mathrm{E}_{\pi(\boldsymbol{\gamma}|\mathcal{D}_{obs})}\left[\mathrm{Dev}_{[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]}(\boldsymbol{\gamma}) \mid \mathcal{D}_{obs}\right] - \mathrm{Dev}_{[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]}(\overline{\boldsymbol{\gamma}})$ being the corresponding effective number of parameters. Also, the DIC for the PDE given the AC and RT is

$$\mathrm{DIC}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} = \mathrm{Dev}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]}(\overline{\boldsymbol{\gamma}}) + 2p_{D[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]}, \tag{4.11}$$

with $p_{D[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} = \mathrm{E}_{\pi(\boldsymbol{\gamma}|\mathcal{D}_{obs})}\left[\mathrm{Dev}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]}(\boldsymbol{\gamma}) \mid \mathcal{D}_{obs}\right] - \mathrm{Dev}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]}(\overline{\boldsymbol{\gamma}})$ being the corresponding effective number of parameters.

With all notations defined above, we propose to decompose the total DIC of the proposed joint model in equation 4.1 into different parts as stated in Corollary 1 by following the idea of D. Zhang et al. (2017) and F. Zhang et al. (2021).

Corollary 1. The DIC and p_D defined in equation 4.1 for the proposed joint model have the following decomposition:

$$\begin{aligned} \mathrm{DIC} &= \mathrm{DIC}_{[\mathrm{RT},\mathrm{PDE}]} + \mathrm{DIC}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} && (4.12)\\ &= \mathrm{DIC}_{[\mathrm{AC},\mathrm{PDE}]} + \mathrm{DIC}_{[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]} && (4.13)\\ &= \mathrm{DIC}_{[\mathrm{AC},\mathrm{RT}]} + \mathrm{DIC}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]}. && (4.14) \end{aligned}$$

Also, the total p_D can decompose as

$$\begin{aligned} p_D &= p_{D[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} + p_{D[\mathrm{RT},\mathrm{PDE}]} && (4.15)\\ &= p_{D[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]} + p_{D[\mathrm{AC},\mathrm{PDE}]} && (4.16)\\ &= p_{D[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} + p_{D[\mathrm{AC},\mathrm{RT}]}. && (4.17) \end{aligned}$$

In addition, we could do a further decomposition of $\mathrm{DIC}_{[\mathrm{RT},\mathrm{PDE}]}$ in equation 4.12 and $\mathrm{DIC}_{[\mathrm{AC},\mathrm{RT}]}$ in equation 4.14. Since the PDE scores follow a normal regression model, we have $[\text{PDE}_i \mid \boldsymbol{\gamma}_{21}] \sim \mathcal{N}(\beta_0, \sigma^2_{\text{PDE}} + \beta_1^2 + \beta_2^2)$ and thus the pdf of PDE_i given $\boldsymbol{\gamma}_{21}$, i.e., $f(\text{PDE}_i \mid \boldsymbol{\gamma}_{21})$, is a normal pdf. Then, the deviance function of the PDE data has the form $\text{Dev}_{[\text{PDE}]}(\boldsymbol{\gamma}_{21}) = -2\sum_{i=1}^{N} \log f(\text{PDE}_i|\boldsymbol{\gamma}_{21})$, and then the DIC of the PDE data is $\text{DIC}_{[\text{PDE}]} = \text{Dev}_{[\text{PDE}]}(\overline{\boldsymbol{\gamma}}_{21}) + 2p_{D[\text{PDE}]}$ and accordingly, $p_{D[\text{PDE}]}$, the effective number of parameters equal to $\text{E}_{\pi(\boldsymbol{\gamma}_{21}|\mathcal{D}_{obs})}[\text{Dev}_{[\text{PDE}]}(\boldsymbol{\gamma}_{21}) \mid \mathcal{D}_{obs}] - \text{Dev}_{[\text{PDE}]}(\overline{\boldsymbol{\gamma}}_{21})$. Similarly, let $f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22})$ and $f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1)$ denote the marginal densities of $\boldsymbol{t}_i$ and $\boldsymbol{y}_i$, respectively, in terms of the joint density of $(\boldsymbol{t}_i, \boldsymbol{y}_i)$, which is $f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1)$ as shown in equation 4.3. Accordingly, we can define the deviance functions for the RT model and the AC model, i.e., we have $\text{Dev}_{[\text{RT}]}(\boldsymbol{\gamma}_{22}) = -2\sum_{i=1}^{N} \log f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22})$, where $f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22})$ is the pdf of a normal distribution $\mathcal{N}(\boldsymbol{\lambda}_{J_i}, \sigma_\tau^2 \mathbf{1}_{J_i}\mathbf{1}'_{J_i} + \bigoplus_{j=1}^{J_i} \sigma_j^2)$, and $\text{Dev}_{[\text{AC}]}(\boldsymbol{\gamma}_1) = -2\sum_{i=1}^{N} \log f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1)$, where $f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1) = \int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i)d\theta_i$. Then, we define $\text{DIC}_{[\text{RT}]} = \text{Dev}_{[\text{RT}]}(\overline{\boldsymbol{\gamma}}_{22}) + 2p_{D[\text{RT}]}$ as the DIC for the RT and $\text{DIC}_{[\text{AC}]} = \text{Dev}_{[\text{AC}]}(\overline{\boldsymbol{\gamma}}_1) + 2p_{D[\text{AC}]}$ as the DIC for the AC. Here, $p_{D[\text{RT}]} = \text{E}_{\pi(\boldsymbol{\gamma}_{22}|\mathcal{D}_{obs})}[\text{Dev}_{[\text{RT}]}(\boldsymbol{\gamma}_{22}) \mid \mathcal{D}_{obs}] - \text{Dev}_{[\text{RT}]}(\overline{\boldsymbol{\gamma}}_{22})$ is the effective number of parameters for the RT model and $p_{D[\text{AC}]} = \text{E}_{\pi(\boldsymbol{\gamma}_1|\mathcal{D}_{obs})}[\text{Dev}_{[\text{AC}]}(\boldsymbol{\gamma}_1) \mid \mathcal{D}_{obs}] - \text{Dev}_{[\text{AC}]}(\overline{\boldsymbol{\gamma}}_1)$ is the effective numbers of parameter for the AC model. Again, although the expectations in $p_{D[\text{PDE}]}$, $p_{D[\text{RT}]}$ and $p_{D[\text{AC}]}$, are taken with respect to $\pi(\boldsymbol{\gamma}_{21} \mid \mathcal{D}_{obs})$, $\pi(\boldsymbol{\gamma}_{22} \mid \mathcal{D}_{obs})$, and $\pi(\boldsymbol{\gamma}_1 \mid \mathcal{D}_{obs})$, respectively, we can still use the MCMC samples drawn from $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$ to estimate those expectations since $\pi(\boldsymbol{\gamma}_{21} \mid \mathcal{D}_{obs})$, $\pi(\boldsymbol{\gamma}_{22} \mid \mathcal{D}_{obs})$, and $\pi(\boldsymbol{\gamma}_1 \mid \mathcal{D}_{obs})$ are the marginal posteriors of $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$.

By the conditional property of the multivariate normal distribution and equation 4.2,

$$[\boldsymbol{t}_i \mid \text{PDE}_i, \boldsymbol{\gamma}_2] \sim \mathcal{N}_{J_i}\left(\boldsymbol{\lambda}_{J_i} - c\sigma_\tau(\text{PDE}_i - \beta_0)\mathbf{1}_{J_i}, \sigma_\tau^2\left[1 - c(\beta_1 \sin\varphi + \beta_2\cos\varphi)\right]\mathbf{1}_{J_i}\mathbf{1}'_{J_i} + \bigoplus_{j=1}^{J_i}\sigma_j^2\right),$$

where $c = (\beta_1 \sin\varphi + \beta_2 \cos\varphi)/(\sigma^2_{\text{PDE}} + \beta_1^2 + \beta_2^2)$. Then, $f(\boldsymbol{t}_i \mid \text{PDE}_i, \boldsymbol{\gamma}_2)$ is also a multivariate normal pdf. According to equation 4.2 and the conditional formula above, there are three cases that the PDE data is independent of the RT data and would not help its analysis: (1) $\sin\varphi = 0$ (i.e., θ_i is not related to the RT model), $\beta_2 = 0$, (2) $\cos\varphi = 0$ (i.e., τ_i is not related to the RT model), $\beta_1 = 0$, and (3) $\beta_1 = 0$ and $\beta_2 = 0$. Scenario (3) is easy to see from the reparameterized PDE model, however, the first two scenarios can be clearly interpreted from Figure 2.1.

Define $\text{DIC}_{[\text{RT}|\text{PDE}]} = \text{Dev}_{[\text{RT}|\text{PDE}]}(\overline{\boldsymbol{\gamma}}_2) + 2p_{D[\text{RT}|\text{PDE}]}$ as the DIC for the RT model given the PDE data, with $\text{Dev}_{[\text{RT}|\text{PDE}]}(\boldsymbol{\gamma}_2) = -2\sum_{i=1}^{N} \log f(\boldsymbol{t}_i \mid \text{PDE}_i, \boldsymbol{\gamma}_2)$ as the deviance function for the RT given the PDE and $p_{D[\text{RT}|\text{PDE}]}$ as the effective number of parameters equal to $\text{E}_{\pi(\boldsymbol{\gamma}_2|\mathcal{D}_{obs})}[\text{Dev}_{[\text{RT}|\text{PDE}]}(\boldsymbol{\gamma}_2) \mid \mathcal{D}_{obs}] - \text{Dev}_{[\text{RT}|\text{PDE}]}(\overline{\boldsymbol{\gamma}}_2)$. Likewise, we could define the DIC for the AC model given the RT data as $\text{DIC}_{[\text{AC}|\text{RT}]} = \text{Dev}_{[\text{AC}|\text{RT}]}(\overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{22}) + 2p_{D[\text{AC}|\text{RT}]}$, where $\text{Dev}_{[\text{AC}|\text{RT}]}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}) = -2\sum_{i=1}^{N} \log f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22})$ with the likelihood $f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22})$ computed through $\int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i) f(\theta_i \mid \boldsymbol{t}_i, \boldsymbol{\gamma}_{22}) d\theta_i$ and the effective number of parameters $p_{D[\text{AC}|\text{RT}]}$ equals to $\text{E}_{\pi(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_{22}|\mathcal{D}_{obs})}[\text{Dev}_{[\text{AC}|\text{RT}]}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}) \mid \mathcal{D}_{obs}] - \text{Dev}_{[\text{AC}|\text{RT}]}(\overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{22})$. In a similar way, we obtain the DIC for the RT model given the item response data using $\text{DIC}_{[\text{RT}|\text{AC}]} = \text{Dev}_{[\text{RT}|\text{AC}]}(\overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{22}) + 2p_{D[\text{RT}|\text{AC}]}$, where $\text{Dev}_{[\text{RT}|\text{AC}]}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22})$ is the deviance function equal to $-2\sum_{i=1}^{N} \log f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22})$ with the likelihood $f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22})$ being $\int f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{22}, \theta_i) f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta_i)\phi(\theta_i) d\theta_i / f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1)$ and the effective number of parameters $p_{D[\text{RT}|\text{AC}]} = \text{E}_{\pi(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_{22}|\mathcal{D}_{obs})}[\text{Dev}_{[\text{RT}|\text{AC}]}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}) \mid \mathcal{D}_{obs}] - \text{Dev}_{[\text{RT}|\text{AC}]}(\overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{22})$. Then, we summarize the further decomposition of the DIC for jointly modeling the RT and the PDE data as well as the RT and item response data in Corollary 2.

Corollary 2. The DIC value of joint modeling of the RT and PDE data as well as the RT and item response data can be decomposed as below

$$\begin{aligned}
&\text{DIC}_{[\text{RT},\text{PDE}]} = \text{DIC}_{[\text{PDE}]} + \text{DIC}_{[\text{RT}|\text{PDE}]},\\
&\text{DIC}_{[\text{AC},\text{RT}]} = \text{DIC}_{[\text{AC}]} + \text{DIC}_{[\text{RT}|\text{AC}]} = \text{DIC}_{[\text{RT}]} + \text{DIC}_{[\text{AC}|\text{RT}]},
\end{aligned} \tag{4.18}$$

and the effective numbers of parameters are decomposed as $p_{D[\text{RT},\text{PDE}]} = p_{D[\text{PDE}]} + p_{D[\text{RT}|\text{PDE}]}$, and $p_{D[\text{AC},\text{RT}]} = p_{D[\text{AC}]} + p_{D[\text{RT}|\text{AC}]} = p_{D[\text{RT}]} + p_{D[\text{AC}|\text{RT}]}$.

By combining Corollary 1 and Corollary 2, we obtain the full decomposition of the DIC under the proposed joint model as shown in Proposition 3.

Proposition 3. The total DIC under the proposed joint model can be decomposed as

$$\mathrm{DIC} = \mathrm{DIC}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} + \mathrm{DIC}_{[\mathrm{RT}|\mathrm{PDE}]} + \mathrm{DIC}_{[\mathrm{PDE}]} \tag{4.19}$$

$$= \mathrm{DIC}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} + \mathrm{DIC}_{[\mathrm{RT}|\mathrm{AC}]} + \mathrm{DIC}_{[\mathrm{AC}]} \tag{4.20}$$

$$= \mathrm{DIC}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} + \mathrm{DIC}_{[\mathrm{AC}|\mathrm{RT}]} + \mathrm{DIC}_{[\mathrm{RT}]}, \tag{4.21}$$

and the corresponding effective number of parameters p_D can be partitioned as

$$p_D = p_{D[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} + p_{D[\mathrm{RT}|\mathrm{PDE}]} + p_{D[\mathrm{PDE}]} \tag{4.22}$$

$$= p_{D[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} + p_{D[\mathrm{RT}|\mathrm{AC}]} + p_{D[\mathrm{AC}]} \tag{4.23}$$

$$= p_{D[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} + p_{D[\mathrm{AC}|\mathrm{RT}]} + p_{D[\mathrm{RT}]}. \tag{4.24}$$

4.1.2. $\Delta\mathrm{DIC}_{\mathrm{AC}}$*,* $\Delta\mathrm{DIC}^*_{\mathrm{AC}}$*,* $\Delta\mathrm{DIC}_{\mathrm{RT}}$*,* $\Delta\mathrm{DIC}^*_{\mathrm{RT}}$ *and* $\Delta\mathrm{DIC}_{\mathrm{PDE}}$

Instead of fitting a joint model, we can separately fit the item response, RTs and PDE scores one by one, by using equation 2.1 for the AC model, equation 2.2 for the RT model with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_j^2)$ and equation 2.3 for the PDE model with $\zeta^* \sim \mathcal{N}(0, \sigma^2_{\mathrm{PDE}})$. Define $\boldsymbol{\gamma}_{\mathrm{PDE}} = (\beta_0, \sigma^2_{\mathrm{PDE}})'$ and $\boldsymbol{\gamma}_{\mathrm{RT}} = (\boldsymbol{\lambda}'_J, \sigma^2_\tau, \sigma^2_1, \cdots, \sigma^2_J)'$; and $\mathcal{D}_{\mathrm{AC},obs} = \{\boldsymbol{y}_i, i = 1, \cdots, N\}$, $\mathcal{D}_{\mathrm{RT},obs} = \{\boldsymbol{t}_i, i = 1, \cdots, N\}$, and $\mathcal{D}_{\mathrm{PDE},obs} = \{\mathrm{PDE}_i, i = 1, \cdots, N\}$ as the observations for each part of multidimensional data. Then, the respective DICs for the AC model, RT and PDE alone are given by

$$\mathrm{DIC}^o_{[\mathrm{AC}]} = \mathrm{Dev}^o_{[\mathrm{AC}]}(\widetilde{\boldsymbol{\gamma}}_1) + 2p^o_{D[\mathrm{AC}]}, \tag{4.25}$$

$$\mathrm{DIC}^o_{[\mathrm{RT}]} = \mathrm{Dev}^o_{[\mathrm{RT}]}(\widetilde{\boldsymbol{\gamma}}_{\mathrm{RT}}) + 2p^o_{D[\mathrm{RT}]}, \tag{4.26}$$

$$\mathrm{DIC}^o_{[\mathrm{PDE}]} = \mathrm{Dev}^o_{[\mathrm{PDE}]}(\widetilde{\boldsymbol{\gamma}}_{\mathrm{PDE}}) + 2p^o_{D[\mathrm{PDE}]}. \tag{4.27}$$

Here, $\mathrm{Dev}^o_{[\mathrm{AC}]}(\boldsymbol{\gamma}_1) = -2\sum\limits_{i=1}^{N} \log f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1) = -2\sum_{i=1}^N \log \int f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta^*_i)\phi(\theta^*_i)d\theta^*_i$ is the deviance function for the AC model alone, with $f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \theta^*_i)$ defined in equation 3.1 by replacing θ_i with θ^*_i. $p^o_{D[\mathrm{AC}]} = \mathrm{E}_{\pi(\boldsymbol{\gamma}_1|\mathcal{D}_{\mathrm{AC},obs})}[\mathrm{Dev}^o_{[\mathrm{AC}]}(\boldsymbol{\gamma}_1) \mid \mathcal{D}_{\mathrm{AC},obs}] - \mathrm{Dev}^o_{[\mathrm{AC}]}(\widetilde{\boldsymbol{\gamma}}_1)$ is the effective number of parameters for the AC model alone, where the expectation is taken respect to the posterior distribution of $\boldsymbol{\gamma}_1$ based on the AC data and $\widetilde{\boldsymbol{\gamma}}_1$ is its posterior mean with $\pi(\boldsymbol{\gamma}_1 \mid \mathcal{D}_{\mathrm{AC},obs}) \propto f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1)\pi(\boldsymbol{\gamma}_1)$ and $\pi(\boldsymbol{\gamma}_1)$ as the prior distribution for $\boldsymbol{\gamma}_1$. Likewise, the deviance function $\mathrm{Dev}^o_{[\mathrm{RT}]}(\boldsymbol{\gamma}_{\mathrm{RT}}) = -2\sum_{i=1}^{N}\log f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{\mathrm{RT}})$ is for the RT model alone, with $f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{\mathrm{RT}}) = \int f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{\mathrm{RT}}, \tau_i^*)\phi(\tau_i^*)d\tau_i^*$ and $f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{\mathrm{RT}}, \tau_i^*)$ is the pdf of the RT model defined in equation 2.2. Further, $p^o_{D[\mathrm{RT}]}$ is the effective number of parameters for the RT model alone, which equals to $\mathrm{E}_{\pi(\boldsymbol{\gamma}_{\mathrm{RT}}\mid\mathcal{D}_{\mathrm{RT},obs})}[\mathrm{Dev}^o_{[\mathrm{RT}]}(\boldsymbol{\gamma}_{\mathrm{RT}}) \mid \mathcal{D}_{\mathrm{RT},obs}] - \mathrm{Dev}^o_{[\mathrm{RT}]}(\widetilde{\boldsymbol{\gamma}}_{\mathrm{RT}})$, with the expectation respect to $\pi(\boldsymbol{\gamma}_{\mathrm{RT}} \mid \mathcal{D}_{\mathrm{RT},obs}) \propto f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{\mathrm{RT}})\pi(\boldsymbol{\gamma}_{\mathrm{RT}})$, $\pi(\boldsymbol{\gamma}_{\mathrm{RT}})$ as the prior for $\boldsymbol{\gamma}_{\mathrm{RT}}$ and $\widetilde{\boldsymbol{\gamma}}_{\mathrm{RT}}$ as its posterior mean. Similarly, for the PDE data alone, $\mathrm{Dev}^o_{[\mathrm{PDE}]}(\boldsymbol{\gamma}_{\mathrm{PDE}})$ is its deviance function, which equals to $-2\sum_{i=1}^{N}\log f(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{\mathrm{PDE}})$ with $f(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{\mathrm{PDE}})$ as the pdf of the normal distribution $\mathcal{N}(\beta_0, \sigma^2_{\mathrm{PDE}})$. The effective number of parameters for the PDE model alone is $p^o_{D[\mathrm{PDE}]} = \mathrm{E}_{\pi(\boldsymbol{\gamma}_{\mathrm{PDE}}\mid\mathcal{D}_{\mathrm{PDE},obs})}[\mathrm{Dev}^o_{[\mathrm{PDE}]}(\boldsymbol{\gamma}_{\mathrm{PDE}}) \mid \mathcal{D}_{\mathrm{PDE},obs}] - \mathrm{Dev}^o_{[\mathrm{PDE}]}(\widetilde{\boldsymbol{\gamma}}_{\mathrm{PDE}})$, with the expectation respect to the posterior distribution of $\boldsymbol{\gamma}_{\mathrm{PDE}}$ given the PDE data alone, i.e., $\pi(\boldsymbol{\gamma}_{\mathrm{PDE}} \mid \mathcal{D}_{\mathrm{PDE},obs}) \propto f(\mathrm{PDE}_i \mid \boldsymbol{\gamma}_{\mathrm{PDE}})\pi(\boldsymbol{\gamma}_{\mathrm{PDE}})$, where $\pi(\boldsymbol{\gamma}_{\mathrm{PDE}})$ is the prior distribution for $\boldsymbol{\gamma}_{\mathrm{PDE}}$.

By comparing the DIC value of the AC model in equation 4.25 with $\mathrm{DIC}_{[\mathrm{AC}\mid\mathrm{RT,PDE}]}$ in equation 4.7 and $\mathrm{DIC}_{[\mathrm{AC}\mid\mathrm{RT}]}$ in equation 4.18, the DIC value of the RT model in equation 4.26 with $\mathrm{DIC}_{[\mathrm{RT}\mid\mathrm{AC,PDE}]}$ in equation 4.10 and $\mathrm{DIC}_{[\mathrm{RT}\mid\mathrm{AC}]}$ in equation 4.18, as well as the DIC value of the PDE model in equation 4.27 with $\mathrm{DIC}_{[\mathrm{PDE}\mid\mathrm{AC,RT}]}$ in equation 4.11, we can determine whether the additional information from the two parts of multidimensional data will help us in modeling the remaining part of the multidimensional data. To be specific, we subtract the difference between the pairs of DIC values, i.e.,

$$\Delta\mathrm{DIC}_{\mathrm{AC}} = \mathrm{DIC}^o_{[\mathrm{AC}]} - \mathrm{DIC}_{[\mathrm{AC}\mid\mathrm{RT,PDE}]}, \tag{4.28}$$

$$\Delta\mathrm{DIC}^*_{\mathrm{AC}} = \mathrm{DIC}^o_{[\mathrm{AC}]} - \mathrm{DIC}_{[\mathrm{AC}\mid\mathrm{RT}]}, \tag{4.29}$$

$$\Delta\mathrm{DIC}_{\mathrm{RT}} = \mathrm{DIC}^o_{[\mathrm{RT}]} - \mathrm{DIC}_{[\mathrm{RT}\mid\mathrm{AC,PDE}]}, \tag{4.30}$$

$$\Delta\mathrm{DIC}^*_{\mathrm{RT}} = \mathrm{DIC}^o_{[\mathrm{RT}]} - \mathrm{DIC}_{[\mathrm{RT}\mid\mathrm{AC}]}, \tag{4.31}$$

$$\Delta\mathrm{DIC}_{\mathrm{PDE}} = \mathrm{DIC}^o_{[\mathrm{PDE}]} - \mathrm{DIC}_{[\mathrm{PDE}\mid\mathrm{AC,RT}]}. \tag{4.32}$$

Here, $\Delta\text{DIC}_{\text{AC}}$ measures the gain in the fit of the item response data by conjointly modeling with the RT and PDE data, $\Delta\text{DIC}_{\text{RT}}$ evaluates the gain in the fit of the RT data via conjointly modeling with the item response and PDE data, and $\Delta\text{DIC}_{\text{PDE}}$ determines the gain in the fit of PDE data through conjointly modeling with the item response and RT data. While, $\Delta\text{DIC}^*_{\text{AC}}$ measures the gain in the fit of the item response data by incorporating only the RT data in the joint modeling, and $\Delta\text{DIC}^*_{\text{RT}}$ calculates the gain in the fit of the RT data by introducing only the item response data in the joint model. Our proposed definitions for $\Delta\text{DIC}_{\text{AC}}$, $\Delta\text{DIC}^*_{\text{AC}}$, $\Delta\text{DIC}_{\text{RT}}$, $\Delta\text{DIC}^*_{\text{RT}}$ and $\Delta\text{DIC}_{\text{PDE}}$ have taken account of a penalty for the additional parameters in the proposed joint model. If the value of $\Delta\text{DIC}_{\text{AC}}$ is large, it implies that by incorporating the information from the RT and PDE data, the proposed joint model indeed helps us obtain a better fit for the item response data. However, $\Delta\text{DIC}_{\text{AC}}$ can be negative, which suggests that fitting a AC model alone might be a better option. Similar interpretations are also applied to $\Delta\text{DIC}_{\text{RT}}$ and $\Delta\text{DIC}_{\text{PDE}}$, $\Delta\text{DIC}^*_{\text{AC}}$ and $\Delta\text{DIC}^*_{\text{RT}}$.

4.2. The Logarithm of the Pseudo-marginal Likelihood Criterion

The Bayes factor (BF), defined as the ratio of the marginal likelihoods of the data under the two competing models, is a fundamental criterion for model comparison under Bayesian inference, which could be viewed as the Bayesian equivalent of the likelihood ratio test. Often, for complex models, computing the marginal likelihoods in BF is not an easy task. In practice, this marginal likelihood can be approximated using the conditional predictive ordinate (CPO), which measures the accuracy of prediction using the idea of the leave-one-out cross-validation (Geisser & Eddy, 1979; Gelfand et al., 1992; Gelfand & Dey, 1994). A summary statistic of CPO is then called the logarithm of the pseudo-marginal likelihood (LPML), which has emerged as an alternative to assess model fit in the field of IRT (Bolt et al., 2012; G. Chen & Luo, 2018; Fujimoto, 2018). In this subsection, we focus on the decomposition of CPO and its summary statistic LPML.

4.2.1. CPO Computation

Let $\mathcal{D}_{obs}^{(-i)} = \{(\boldsymbol{y}_k', \boldsymbol{t}_k', \text{PDE}_k)', k = 1, \cdots, i-1, i+1, \cdots, N\}$ denote the observed data with the ith participant deleted. Then, for the ith participant, the corresponding conditional predictive ordinate (CPO) is defined through the posterior predictive density of $(\boldsymbol{y}_i', \boldsymbol{t}_i', \text{PDE}_i)'$, that is,

$$\text{CPO}_i = \pi(\boldsymbol{y}_i, \boldsymbol{t}_i, \text{PDE}_i \mid \mathcal{D}_{obs}^{(-i)}) = \int f(\boldsymbol{y}_i, \boldsymbol{t}_i, \text{PDE}_i \mid \boldsymbol{\gamma})\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}^{(-i)})d\boldsymbol{\gamma}, \tag{4.33}$$

where $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}^{(-i)})$ is computed via $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}^{(-i)}) = \prod_{k=1,k\neq i}^{N} f(\boldsymbol{y}_k, \boldsymbol{t}_k, \text{PDE}_k \mid \boldsymbol{\gamma})\pi(\boldsymbol{\gamma})/c(\mathcal{D}_{obs}^{(-i)})$ and $c(\mathcal{D}_{obs}^{(-i)}) = \int \prod_{k=1,k\neq i}^{N} f(\boldsymbol{y}_k, \boldsymbol{t}_k, \text{PDE}_k \mid \boldsymbol{\gamma})\pi(\boldsymbol{\gamma})d\boldsymbol{\gamma}$ is the normalizing constant. Following M.-H. Chen et al. (2000), CPO_i in equation 4.33 can be rewritten as

$$\text{CPO}_i = \left[\int \left[f(\boldsymbol{y}_i, \boldsymbol{t}_i, \text{PDE}_i \mid \boldsymbol{\gamma})\right]^{-1} \pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})d\boldsymbol{\gamma}\right]^{-1}, \tag{4.34}$$

which makes the Monte Carlo estimate of CPO_i much easier by drawing MCMC samples from the posterior distribution given $\mathcal{D}_{obs}$ instead of $\mathcal{D}_{obs}^{(-i)}$. The expression of CPO_i given in equation 4.34 is also called the CPO Identity I as discussed in D. Zhang et al. (2017). Assuming that M is the total number of iterations in the MCMC sample and $\{\boldsymbol{\gamma}^{[m]}, m = 1, \cdots, M\}$ are the M values of $\boldsymbol{\gamma}$ drawn from $\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})$, a Monte Carlo estimate of CPO_i is given by

$$\widehat{\text{CPO}}_i = M \Big/ \sum_{m=1}^{M} \left[f(\boldsymbol{y}_i, \boldsymbol{t}_i, \text{PDE}_i \mid \boldsymbol{\gamma}^{[m]})\right]^{-1}. \tag{4.35}$$

As discussed in Section 3, $f(\boldsymbol{y}_i, \boldsymbol{t}_i, \text{PDE}_i \mid \boldsymbol{\gamma}^{[m]})$ depends only on a one-dimensional integral (see equation 3.4 for more details), which makes the use of the CPO Identity I possible for computing $\widehat{\text{CPO}}_i$. Similarly, we can define the item response data alone without the ith participant as $\mathcal{D}_{\text{AC},obs}^{(-i)} = \{\boldsymbol{y}_k, k = 1, \cdots, i-1, i+1, \cdots, N\}$, the RT data alone without the ith participant as $\mathcal{D}_{\text{RT},obs}^{(-i)} = \{\boldsymbol{t}_k, k = 1, \cdots, i-1, i+1, \cdots, N\}$ and the PDE data alone without the ith participant as $\mathcal{D}^{(-i)}_{\text{PDE},obs} = \{\text{PDE}_k, k = 1, \cdots, i-1, i+1, \cdots, N\}$. Then, $\text{CPO}_{i,\text{AC}} = \left[\int [f(\boldsymbol{y}_i \mid \gamma_1)]^{-1} \pi(\boldsymbol{\gamma}_1 \mid \mathcal{D}_{\text{AC},obs}) d\boldsymbol{\gamma}_1\right]^{-1}$ is the CPO for the ith participant with respect to the AC alone, $\text{CPO}_{i,\text{RT}} = \left[\int [f(\boldsymbol{t}_i \mid \boldsymbol{\gamma}_{\text{RT}})]^{-1} \pi(\boldsymbol{\gamma}_{\text{RT}} \mid \mathcal{D}_{\text{RT},obs}) d\boldsymbol{\gamma}_{\text{RT}}\right]^{-1}$ is the CPO of the ith participant for the RT alone, and the CPO of the ith participant for the PDE model alone is $\text{CPO}_{i,\text{PDE}} = \left[\int [f(\text{PDE}_i \mid \boldsymbol{\gamma}_{\text{PDE}})]^{-1} \pi(\boldsymbol{\gamma}_{\text{PDE}} \mid \mathcal{D}_{\text{PDE},obs}) d\boldsymbol{\gamma}_{\text{PDE}}\right]^{-1}$. Then we can estimate them using

$$\widehat{\text{CPO}}_{i,\text{AC}} = M \Big/ \sum_{m=1}^{M} \left[f(\boldsymbol{y}_i \mid \widetilde{\boldsymbol{\gamma}}_1^{[m]})\right]^{-1}, \tag{4.36}$$

$$\widehat{\text{CPO}}_{i,\text{RT}} = M \Big/ \sum_{m=1}^{M} \left[f(\boldsymbol{t}_i \mid \widetilde{\boldsymbol{\gamma}}_{\text{RT}}^{[m]})\right]^{-1}, \tag{4.37}$$

$$\widehat{\text{CPO}}_{i,\text{PDE}} = M \Big/ \sum_{m=1}^{M} \left[f(\text{PDE}_i \mid \widetilde{\boldsymbol{\gamma}}_{\text{PDE}}^{[m]})\right]^{-1} \tag{4.38}$$

with $\{\widetilde{\boldsymbol{\gamma}}_1^{[m]}, m = 1, \cdots, M\}$ drawn from the posterior distribution $\pi(\boldsymbol{\gamma}_1 \mid \mathcal{D}_{\text{AC},obs})$, $\{\widetilde{\boldsymbol{\gamma}}_{\text{RT}}^{[m]}, m = 1, \cdots, M\}$ drawn from the posterior distribution $\pi(\boldsymbol{\gamma}_{\text{RT}} \mid \mathcal{D}_{\text{RT},obs})$, and $\{\widetilde{\boldsymbol{\gamma}}_{\text{PDE}}^{[m]}, m = 1, \cdots, M\}$ drawn from the posterior distribution $\pi(\boldsymbol{\gamma}_{\text{PDE}} \mid \mathcal{D}_{\text{PDE},obs})$.

4.2.2. CPO Decomposition

We are going to introduce three ways of decomposition for the CPO to facilitate our comparisons below, which include 1) the AC model alone versus the AC model given the additional information from the RT and PDE data, 2) the RT model alone versus the RT model given the additional information from the item response and PDE data, and 3) the PDE model alone versus the PDE model given the additional information from the RT and item response data. From equation 4.33, we can also write CPO_i as

$$\text{CPO}_i = \frac{c(\mathcal{D}_{obs})}{c(\mathcal{D}^{(-i)}_{obs})} = \frac{f(\boldsymbol{y}_i, \boldsymbol{t}_i, \text{PDE}_i \mid \boldsymbol{\gamma}) \pi(\boldsymbol{\gamma} | \mathcal{D}^{(-i)}_{obs})}{\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs})}. \tag{4.39}$$

Notice that the joint pdf of $(\boldsymbol{y}_i', \boldsymbol{t}_i', \mathrm{PDE}_i)'$ can be written as the product of the conditional pdf and the marginal pdf in three different ways as follows

$$f(\boldsymbol{y}_i, \boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}) = f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) f(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2) \tag{4.40}$$

$$= f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1) \tag{4.41}$$

$$= f(\mathrm{PDE}_i \mid \boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1), \tag{4.42}$$

Here, $f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, $f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and $f(\mathrm{PDE}_i \mid \boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ are given in equation 4.6, equation 4.8, and equation 4.9, respectively; $f(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2)$ is a multivariate normal pdf derived from equation 4.2, while $f(\mathrm{PDE}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{21}, \boldsymbol{\gamma}_1)$ and $f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_{22}, \boldsymbol{\gamma}_1)$ are defined in equation 4.4 and equation 4.3, respectively. Further, we can partition the posterior distributions of $\boldsymbol{\gamma}$ given $\mathcal{D}_{obs}^{(-i)}$ and $\mathcal{D}_{obs}$, respectively, as

$$\pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}^{(-i)}) = \pi(\boldsymbol{\gamma}_2 \mid \mathcal{D}_{obs}^{(-i)}) \pi(\boldsymbol{\gamma}_1 \mid \boldsymbol{\gamma}_2, \mathcal{D}_{obs}^{(-i)}) \tag{4.43}$$

$$= \pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21} \mid \mathcal{D}_{obs}^{(-i)}) \pi(\boldsymbol{\gamma}_{22} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21}, \mathcal{D}_{obs}^{(-i)}) \tag{4.44}$$

$$= \pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22} \mid \mathcal{D}_{obs}^{(-i)}) \pi(\boldsymbol{\gamma}_{21} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}, \mathcal{D}_{obs}^{(-i)}), \tag{4.45}$$

$$\text{and} \quad \pi(\boldsymbol{\gamma} \mid \mathcal{D}_{obs}) = \pi(\boldsymbol{\gamma}_2 \mid \mathcal{D}_{obs}) \pi(\boldsymbol{\gamma}_1 \mid \boldsymbol{\gamma}_2, \mathcal{D}_{obs}) \tag{4.46}$$

$$= \pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21} \mid \mathcal{D}_{obs}) \pi(\boldsymbol{\gamma}_{22} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21}, \mathcal{D}_{obs}) \tag{4.47}$$

$$= \pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22} \mid \mathcal{D}_{obs}) \pi(\boldsymbol{\gamma}_{21} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}, \mathcal{D}_{obs}). \tag{4.48}$$

Using equation 4.39 together with equation 4.40-equation 4.48, we can decompose the value of CPO for the ith participant in three different ways as

$$\mathrm{CPO}_i = \mathrm{CPO}_{i,[\mathrm{RT,PDE}]} \mathrm{CPO}_{i,[\mathrm{AC|RT,PDE}]} \tag{4.49}$$

$$= \mathrm{CPO}_{i,[\mathrm{AC,RT}]} \mathrm{CPO}_{i,[\mathrm{PDE|AC,RT}]} \tag{4.50}$$

$$= \mathrm{CPO}_{i,[\mathrm{AC,PDE}]} \mathrm{CPO}_{i,[\mathrm{RT|AC,PDE}]}. \tag{4.51}$$

Here, $\mathrm{CPO}_{i,[\mathrm{RT,PDE}]} = f(\boldsymbol{t}_i, \mathrm{PDE}_i \mid \boldsymbol{\gamma}_2) \pi(\boldsymbol{\gamma}_2 \mid \mathcal{D}_{obs}^{(-i)}) / \pi(\boldsymbol{\gamma}_2 \mid \mathcal{D}_{obs})$, by analogy, can be viewed as the CPO of the ith participant for jointly modeling the RT and PDE data; $\mathrm{CPO}_{i,[\mathrm{AC,RT}]}$

can similarly be viewed as the CPO of the ith participant for jointly modeling the item response and RT data, which equals to $f(\boldsymbol{t}_i, \boldsymbol{y}_i \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22})\pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22} \mid \mathcal{D}_{obs}^{(-i)})/\pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22} \mid \mathcal{D}_{obs})$; and $\text{CPO}_{i,[\text{AC,PDE}]} = f(\boldsymbol{y}_i, \text{PDE}_i \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21})\pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21} \mid \mathcal{D}_{obs}^{(-i)})/\pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21} \mid \mathcal{D}_{obs})$ can be interpreted as the CPO of the ith participant for jointly modeling the item response and PDE data. While, $\text{CPO}_{i,[\text{AC}|\text{RT,PDE}]}$, $\text{CPO}_{i,[\text{PDE}|\text{AC,RT}]}$ and $\text{CPO}_{i,[\text{RT}|\text{AC,PDE}]}$ are our major focuses in the CPO decomposition, and it is not hard to derive their corresponding formula, where $\text{CPO}_{i,[\text{AC}|\text{RT,PDE}]} = f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \text{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\pi(\boldsymbol{\gamma}_1 \mid \boldsymbol{\gamma}_2, \mathcal{D}_{obs}^{(-i)})/\pi(\boldsymbol{\gamma}_1 \mid \boldsymbol{\gamma}_2, \mathcal{D}_{obs})$ is regarded as the CPO of the item responses given the additional information from the RT and PDE data for the ith participant; $\text{CPO}_{i,[\text{PDE}|\text{AC,RT}]}$, interpreted as the CPO of the PDE data given the additional information from the item response and RT data for the ith participant, equals to $f(\text{PDE}_i \mid \boldsymbol{t}_i, \boldsymbol{y}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\pi(\boldsymbol{\gamma}_{21} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}, \mathcal{D}_{obs}^{(-i)})/\pi(\boldsymbol{\gamma}_{21} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}, \mathcal{D}_{obs})$; $\text{CPO}_{i,[\text{RT}|\text{AC,PDE}]} = f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \text{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\pi(\boldsymbol{\gamma}_{22} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21}, \mathcal{D}_{obs}^{(-i)})/\pi(\boldsymbol{\gamma}_{22} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21}, \mathcal{D}_{obs})$ is the CPO of the RT data given the additional information from the item response and PDE data for the ith participant. With some algebra, we can rewrite conditional CPOs via the following identities:

$$\text{CPO}_{i,[\text{AC}|\text{RT,PDE}]} = \left[\int \left[f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \text{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\right]^{-1} \pi(\boldsymbol{\gamma}_1 \mid \boldsymbol{\gamma}_2, \mathcal{D}_{obs}) d\boldsymbol{\gamma}_1\right]^{-1}, \tag{4.52}$$

$$\text{CPO}_{i,[\text{PDE}|\text{AC,RT}]} = \left[\int \left[f(\text{PDE}_i \mid \boldsymbol{t}_i, \boldsymbol{y}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\right]^{-1} \pi(\boldsymbol{\gamma}_{21} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{22}, \mathcal{D}_{obs}) d\boldsymbol{\gamma}_{21}\right]^{-1}, \tag{4.53}$$

$$\text{CPO}_{i,[\text{RT}|\text{AC,PDE}]} = \left[\int \left[f(\boldsymbol{t}_i \mid \boldsymbol{y}_i, \text{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\right]^{-1} \pi(\boldsymbol{\gamma}_{22} \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_{21}, \mathcal{D}_{obs}) d\boldsymbol{\gamma}_{22}\right]^{-1}, \tag{4.54}$$

which imply that we can mimic the idea of Monte Carlo estimation of CPO_i in equation 4.35 to estimate $\text{CPO}_{i,[\text{AC}|\text{RT,PDE}]}$, $\text{CPO}_{i,[\text{PDE}|\text{AC,RT}]}$, and $\text{CPO}_{i,[\text{RT}|\text{AC,PDE}]}$. Some similar identities can also be derived for $\text{CPO}_{i,[\text{RT,PDE}]}$, $\text{CPO}_{i,[\text{AC,PDE}]}$, and $\text{CPO}_{i,[\text{AC,RT}]}$, however, since those are not our major focus, we omit the details.

To show why equation 4.52 holds, first, we manipulate relationships between the joint and conditional distributions, which yields

$$\pi(\boldsymbol{\gamma}_2 \mid D_{obs}^{-i})/\pi(\boldsymbol{\gamma}_2 \mid \mathcal{D}_{obs}) = \text{CPO}_i \int \left[f(\boldsymbol{y}_i, \boldsymbol{t}_i, \text{PDE}_i \mid \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\right]^{-1} \pi(\boldsymbol{\gamma}_1 \mid \boldsymbol{\gamma}_2, \mathcal{D}_{obs}) d\boldsymbol{\gamma}_1.$$

Then, by plugging this equation above into the expression of $\mathrm{CPO}_{i,[\mathrm{RT},\mathrm{PDE}]}$, we obtain that $\mathrm{CPO}_{i,[\mathrm{RT},\mathrm{PDE}]} = \mathrm{CPO}_i \int \left[f(\boldsymbol{y}_i \mid \boldsymbol{t}_i, \mathrm{PDE}_i, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2,) \right]^{-1} \pi(\boldsymbol{\gamma}_1 \mid \boldsymbol{\gamma}_2, \mathcal{D}_{obs}) d\boldsymbol{\gamma}_1$ and then comparing $\mathrm{CPO}_{i,[\mathrm{RT},\mathrm{PDE}]}$ derived here with equation 4.49, we can conclude equation 4.52 holds. It is clearly seen that if only $\mathrm{CPO}_{i,[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]}$ is of interest, it is not necessary to compute the overall CPO_i. Similar discussions are also applied to equation 4.53 and equation 4.54.

As $\boldsymbol{\gamma}_2$ is unknown, following the discussion of D. Zhang et al. (2017), we plug the posterior mean of $\boldsymbol{\gamma}_2$ (i.e., $\overline{\boldsymbol{\gamma}}_2$) into equation 4.52 and then assuming $\{\boldsymbol{\gamma}_1^{[m]}, m = 1, \cdots, M\}$ is drawn from $\pi(\boldsymbol{\gamma}_1 \mid \overline{\boldsymbol{\gamma}}_2, \mathcal{D}_{obs})$, we have

$$\widehat{\mathrm{CPO}}_{i,[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} = M \Big/ \sum_{m=1}^{M} \left[f(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_1^{[m]}, \overline{\boldsymbol{\gamma}}_2, \boldsymbol{t}_i, \mathrm{PDE}_i) \right]^{-1}. \tag{4.55}$$

Similarly, we also have

$$\widehat{\mathrm{CPO}}_{i,[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} = M \Big/ \sum_{m=1}^{M} \left[f(\mathrm{PDE}_i \mid \overline{\boldsymbol{\gamma}}_1, \boldsymbol{\gamma}_{21}^{[m]}, \overline{\boldsymbol{\gamma}}_{22}, \boldsymbol{t}_i, \boldsymbol{y}_i) \right]^{-1}, \tag{4.56}$$

$$\widehat{\mathrm{CPO}}_{i,[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]} = M \Big/ \sum_{m=1}^{M} \left[f(\boldsymbol{t}_i \mid \overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{21}, \boldsymbol{\gamma}_{22}^{[m]}, \boldsymbol{y}_i, \mathrm{PDE}_i) \right]^{-1} \tag{4.57}$$

with $\{\boldsymbol{\gamma}_{21}^{[m]}, m = 1, \cdots, M\}$ is drawn from $\pi(\boldsymbol{\gamma}_{21} \mid \overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{22}, \mathcal{D}_{obs})$ and $\{\boldsymbol{\gamma}_{22}^{[m]}, m = 1, \cdots, M\}$ is drawn from $\pi(\boldsymbol{\gamma}_{22} \mid \overline{\boldsymbol{\gamma}}_1, \overline{\boldsymbol{\gamma}}_{21}, \mathcal{D}_{obs})$.

4.2.3. LPML and LPML Decomposition

Following Ibrahim et al. (2001), we define the LPML of the proposed joint model as $\mathrm{LPML} = \sum_{i=1}^{N} \log \widehat{\mathrm{CPO}}_i$, where $\widehat{\mathrm{CPO}}_i$ is computed using equation 4.35. Further, it is easy to see that the LPML of the proposed joint model can be decomposed as

$$\mathrm{LPML} = \mathrm{LPML}_{[\mathrm{RT},\mathrm{PDE}]} + \mathrm{LPML}_{[\mathrm{AC}|\mathrm{RT},\mathrm{PDE}]} \tag{4.58}$$

$$= \mathrm{LPML}_{[\mathrm{AC},\mathrm{RT}]} + \mathrm{LPML}_{[\mathrm{PDE}|\mathrm{AC},\mathrm{RT}]} \tag{4.59}$$

$$= \mathrm{LPML}_{[\mathrm{AC},\mathrm{PDE}]} + \mathrm{LPML}_{[\mathrm{RT}|\mathrm{AC},\mathrm{PDE}]}, \tag{4.60}$$

by using equation 4.49, equation 4.50, equation 4.51 and the property of the logarithm that the logarithm of a product is the sum of the logarithms of the factors. Here, $\text{LPML}_{[\text{RT,PDE}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,[\text{RT,PDE}]}$, where $\widehat{\text{CPO}}_{i,[\text{RT,PDE}]}$ is the estimate of $\text{CPO}_{i,[\text{RT,PDE}]}$; $\text{LPML}_{[\text{AC,RT}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,[\text{AC,RT}]}$ with $\widehat{\text{CPO}}_{i,[\text{AC,RT}]}$ being the estimate of $\text{CPO}_{i,[\text{AC,RT}]}$; and $\text{LPML}_{[\text{AC,PDE}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,[\text{AC,PDE}]}$ with $\widehat{\text{CPO}}_{i,[\text{AC,PDE}]}$ being the estimate of $\text{CPO}_{i,[\text{AC,PDE}]}$. We also have $\text{LPML}_{[\text{AC}|\text{RT,PDE}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,[\text{AC}|\text{RT,PDE}]}$, where $\widehat{\text{CPO}}_{i,[\text{AC}|\text{RT,PDE}]}$ is estimated by equation 4.55; $\text{LPML}_{[\text{PDE}|\text{AC,RT}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,[\text{PDE}|\text{AC,RT}]}$ with $\widehat{\text{CPO}}_{i,[\text{PDE}|\text{AC,RT}]}$ estimated by equation 4.56; and let $\widehat{\text{CPO}}_{i,[\text{RT}|\text{AC,PDE}]}$ is estimated by equations 4.55, then $\text{LPML}_{[\text{RT}|\text{AC,PDE}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,[\text{RT}|\text{AC,PDE}]}$.

4.2.4. $\Delta\text{LPML}_{\text{AC}}$*,* $\Delta\text{LPML}_{\text{PDE}}$ *and* $\Delta\text{LPML}_{\text{RT}}$

Since a larger value of LPML suggests a better model fit, by analogy to the derivations of $\Delta\text{DIC}_{\text{AC}}$, $\Delta\text{DIC}_{\text{PDE}}$, and $\Delta\text{DIC}_{\text{RT}}$, we can define the differences of LPML as

$$\Delta\text{LPML}_{\text{AC}} = \text{LPML}_{[\text{AC}|\text{RT,PDE}]} - \text{LPML}_{[\text{AC}]}, \tag{4.61}$$

$$\Delta\text{LPML}_{\text{RT}} = \text{LPML}_{[\text{RT}|\text{AC,PDE}]} - \text{LPML}_{[\text{RT}]}, \tag{4.62}$$

$$\Delta\text{LPML}_{\text{PDE}} = \text{LPML}_{[\text{PDE}|\text{AC,RT}]} - \text{LPML}_{[\text{PDE}]}, \tag{4.63}$$

where $\text{LPML}_{[\text{AC}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,\text{AC}}$, $\text{LPML}_{[\text{RT}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,\text{RT}}$, and $\text{LPML}_{[\text{PDE}]} = \sum_{i=1}^{N} \log \widehat{\text{CPO}}_{i,\text{PDE}}$ with $\widehat{\text{CPO}}_{i,\text{AC}}$, $\widehat{\text{CPO}}_{i,\text{RT}}$, and $\widehat{\text{CPO}}_{i,\text{PDE}}$ estimated through equations 4.36, 4.37, and 4.38, respectively. To save the space in this paper, we haven't reported $\Delta\text{LPML}^*_{\text{AC}}$ (analogy to $\Delta\text{DIC}^*_{\text{AC}}$) to quantify the gain in the fit of the item response data by only conjointly model with the RT data, as well as $\Delta\text{LPML}^*_{\text{RT}}$ (analogy to $\Delta\text{DIC}^*_{\text{RT}}$) to measure the gain in the fit of the RT data by introducing only the item response data in the joint model. The formula for $\Delta\text{LPML}^*_{\text{AC}}$ and $\Delta\text{LPML}^*_{\text{RT}}$ are similar to those in equation 4.61 and equation 4.62.

As an example, if we take an exponential on both sides of equation 4.61, the right side will become a ratio of two pseudo likelihood functions, which compares the likelihood function of the item response data given additional information from the RT and PDE data to the likelihood function of item response data alone. Thus, we can view $\exp(\Delta\text{LPML}_{\text{AC}})$, (similarly, for $\exp(\Delta\text{LPML}_{\text{RT}})$ and $\exp(\Delta\text{LPML}_{\text{PDE}})$) as a pseduo-BF, and use it to quantify the gain in the fit of the item response data (similarly, for the RT data or the PDE data) due to the joint model. A model with a large value of ΔLPML (corresponding to a large pseudo-BF value) means the joint model is more favorable. Following the rule of thumb for the scale of BF suggested by Jeffreys (1961) or Kass and Raftery (1995), we can determine which models we are going to support.

5. Simulation Study

In this section, we first investigate the performance of the MCMC methods in recovery of the true parameters for the joint model using the simulation data. Next, we evaluate the empirical performance of model assessment criteria, including $\Delta\text{DIC}_{\text{AC}}$, $\Delta\text{DIC}^*_{\text{AC}}$, $\Delta\text{DIC}_{\text{RT}}$, $\Delta\text{DIC}^*_{\text{RT}}$, $\Delta\text{DIC}_{\text{PDE}}$, $\Delta\text{LPML}_{\text{AC}}$, $\Delta\text{LPML}_{\text{RT}}$, and $\Delta\text{LPML}_{\text{PDE}}$, over different numbers of items and participants.

5.1. The Recovery Study of Parameters in the Model

We generate the simulated data from the reparameterized and reformulated joint model proposed in equation 2.7 to equation 2.8 of Section 2, which is equivalent to generate the data directly from equation 2.1 to equation 2.4. The true parameters in the model are set in a way to resemble the posterior means from the empirical data as follows. First, we draw item discrimination parameters a_js from a uniform distribution $U(0.3, 1.9)$, item difficulty parameters b_js from $\mathcal{N}(0, 1)$ and item intensity parameters λ_js from $U(0.8, 1.9)$. Also, we assume the variances of ϵ_{ij}s in the RT model, i.e, σ_j^2s are sampled from a uniform distribution $U(0.14, 0.75)$. For the latent attribute parameters θ_is and τ_is at the participant level, we assume both are drawn from a standard normal distribution $\mathcal{N}(0,1)$. We set the coefficients $\beta_0, \beta_1, \beta_2$ and σ^2_{PDE} in equation 2.9 to be 93.8, 14.5, 3.5, and 57, respectively. Further, let the correlation related parameter φ in equation 2.8 be 0.411 and the standard deviation parameter of speed σ_τ in equation 2.8 be 0.367. Notice these settings are corresponding to using $\sigma_{12} = 0.147$, $\sigma_{13} = \beta_1 = 14.5$, $\sigma^2_\tau = 0.135$, $\sigma_{23} = 3.30$ and $\sigma_{33} = 279.5$ in equation 2.4 under the original joint model.

For the number of participants, we consider two cases, $N = 125$ and $N = 250$, and for the number of items in a test, we also consider two cases, $J = 20$ and $J = 40$. Hence, there are a total of $2 \times 2 = 4$ different scenarios. For the scenario of $J = 20$ items, we use the same values of item parameters selected from the scenario of $J = 40$ items. Similarly, for the case of $N = 125$ participants, we use the same values of the latent attribute parameters selected from the case of $N = 250$ respondents. For each scenario, we independently simulate 500 replications of the data. The running time depends on the numbers of participants and items as well as the length of the MCMC chain. In the simulation scenario of $N = 125$ and $J = 20$, for each simulated data set, it takes 3.9 minutes for run a 25,000 MCMC iterations in R codes on an Intel i7 processor machine with 16 GB of RAM memory using a Windows 10 operating system. To compute all of the model assessment criteria based on the decomposition of DIC, it takes about 22 minutes to run the Fortran 95 codes on an Intel i686 processor machine with 16 GB of RAM memory using a GNU/Linux operating system. The running time increases proportionally when the numbers of individuals and items increase.

For each replication of the data, using 25,000 MCMC iterations with a burn-in period of 5000 iterations and thinning the sample for every two iterations, we compute the posterior mean and 95% highest posterior density (HPD) intervals (M.-H. Chen & Shao, 1999) for each parameter. Then, we compare the posterior means of the parameters relative to their true values among 500 replications. By averaging these values, we obtain the bias using Bias $= \sum_{r=1}^{R} (\widehat{\vartheta}_r - \vartheta)/R$ and the mean squared error (MSE) through

$\text{MSE} = \sum_{r=1}^{R} (\widehat{\vartheta}_r - \vartheta)^2/R$, where ϑ is the parameter of interest, $\widehat{\vartheta}_r$ is the posterior mean of the parameter in the rth replication, with $r = 1, \cdots, R$ and $R = 500$. In addition, we compute the frequentist coverage probability (CP) by counting among 500 replications how many the 95% HPD intervals of the parameter will contain the true value of the parameter. Further, we report the sample standard deviation (SD) of the posterior estimates over replica datasets, i.e., $\text{SD} = \sum_{r=1}^{R} \sqrt{\sum_{m=1}^{M} (\vartheta_r^{[m]} - \widehat{\vartheta}_r)^2/(M-1)}/R$, where $\vartheta_r^{[m]}$ is the mth iteration of the MCMC sample for the parameter ϑ_r in the rth replication of the data and $m = 1, \cdots, M$; and we compute the simulation standard error (SE) via
$\text{SE} = \sqrt{\sum_{r=1}^{R} (\widehat{\vartheta}_r - \frac{1}{R} \sum_{\ell=1}^{R} \widehat{\vartheta}_\ell)^2/(R-1)}$.

Table 5.1 shows a summary of the simulation results for $\boldsymbol{a} = (a_1, \cdots, a_J)'$, $\boldsymbol{b} = (b_1, \cdots, b_J)'$ and $\boldsymbol{\lambda}_J = (\lambda_1, \cdots, \lambda_J)'$, respectively. Due to a large number of items, in Table 5.1, the content in the intersection of Column 'Bias' and Row '$\boldsymbol{a}$ (Min, Max)' presents the median value of the bias among all a_js and the bracket shows the range of the bias values for all a_js. Similar interpretation can be applied to Bias, MSE, CP and SD for all other columns and rows in the table. In analogy, we summarize the latent attribute parameters of respondents by grouping them as $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ shown in Table 5.2, where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_N)'$ and $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)'$. For example, the content in Column of 'Bias' intersected with Row '$\boldsymbol{\theta}$ (Min, Max)' provides the median value of the bias among all θ_is and the bracket there is the range of the bias for all θ_is. The rest columns and rows of Bias, MSE, CP and SD can be interpreted accordingly in Table 5.2. To save space, we put the SE results of all unknown parameters in Table S.1 of the Supplementary Materials.

Based on these tables, we can draw some conclusions for the recovery performance of item and latent attribute parameters. First, the values of Bias and MSE for the items parameters decrease substantially as the number of participants increases from $N = 125$ to $N = 250$. In most cases of $N = 250$, the median values of MSE for the item parameters are below 0.05. Second, the values of Bias and MSE for latent attribute parameters have

Table 5.1: Summary of item parameters in the simulation

N	J		Bias	SD	MSE	CP(%)
125	20	$\boldsymbol{a}$ (Min, Max)	-.064 (-.296, .101)	.267 (.175, .401)	.069 (.029, .162)	.936 (.866, .984)
		$\boldsymbol{b}$ (Min, Max)	.108 (-.171, .183)	.298 (.186, .547)	.091 (.031, .155)	.967 (.926, 1.000)
		$\boldsymbol{\lambda}$ (Min, Max)	.011 (.006, .017)	.066 (.050, .084)	.003 (.002, .006)	.976 (.956, .988)
		β_0	-1.740	1.483	3.502	.940
		β_1	-.230	1.165	.755	.986
		β_2	-.109	1.050	1.153	.944
		σ^2_{PDE}	-1.382	12.850	178.797	.930
	40	$\boldsymbol{a}$ (Min, Max)	-.118 (-.315, .213)	.262 (.161, .392)	.076 (.022 , .175)	.920 (.838, .994)
		$\boldsymbol{b}$ (Min, Max)	.154 (-.221, .282)	.271 (.178, .557)	.097 (.023, .177)	.963 (.900, .992)
		$\boldsymbol{\lambda}$ (Min, Max)	.008 (.006, .013)	.065 (.049, .084)	.003 (.001, .006)	.978 (.958, .998)
		β_0	-1.625	1.425	3.088	.964
		β_1	-.992	1.010	1.458	.926
		β_2	-.021	.885	.720	.968
		σ^2_{PDE}	-1.348	10.067	107.075	.950
250	20	$\boldsymbol{a}$ (Min, Max)	-.014 (-.120, .075)	.198 (.133, .329)	.037 (.019, .090)	.954 (.930, .982)
		$\boldsymbol{b}$ (Min, Max)	.091 (-.058, .121)	.199 (.118, .441)	.037 (.014, .125)	.951 (.926, .990)
		$\boldsymbol{\lambda}$ (Min, Max)	-.002 (-.005, .001)	.046 (.035, .059)	.002 (.001, .003)	.978 (.960, .988)
		β_0	-1.220	1.049	1.708	.968
		β_1	-.173	.842	.335	.998
		β_2	.086	.885	.452	.960
		σ^2_{PDE}	.296	8.794	81.879	.942
	40	$\boldsymbol{a}$ (Min, Max)	-.041 (-.130, .171)	.201 (.123, .320)	.038 (.012, .087)	.944 (.912, .976)
		$\boldsymbol{b}$ (Min, Max)	.105 (-.053, .251)	.177 (.116, .457)	.037 (.013, .151)	.955 (.898, .984)
		$\boldsymbol{\lambda}$ (Min, Max)	-.002 (-.006, .001)	.046 (.034, .059)	.002 (.001, .003)	.976 (.958, .992)
		β_0	-1.198	1.023	1.678	.942
		β_1	-.605	.757	.609	.966
		β_2	.103	.608	.373	.950
		σ^2_{PDE}	.259	6.856	46.750	.936

Table 5.2: Summary of latent attribute parameters for respondents in the simulation

J	N		Bias	SD	MSE	CP(%)
20	125	$\boldsymbol{\theta}$ (Min, Max)	.103 (-.337, .510)	.355 (.341, .430)	.120 (.095, .370)	.954 (.810, .984)
		$\boldsymbol{\tau}$ (Min, Max)	-.025 (-.372, .386)	.406 (.396, .473)	.137 (.107, .263)	.966 (.892, .996)
	250	$\boldsymbol{\theta}$ (Min, Max)	.080 (-.379, .554)	.341 (.329, .429)	.111 (.082, .434)	.954 (.776, .984)
		$\boldsymbol{\tau}$ (Min, Max)	-.035 (-.528, .500)	.380 (.374, .420)	.135 (.097, .389)	.956 (.776, .986)
40	125	$\boldsymbol{\theta}$ (Min, Max)	.095 (-.070, .344)	.299 (.282, .398)	.085 (.066, .208)	.958 (.876, .992)
		$\boldsymbol{\tau}$ (Min, Max)	-.024 (-.243, .209)	.312 (.298, .395)	.077 (.064, .127)	.970 (.934, .996)
	250	$\boldsymbol{\theta}$ (Min, Max)	.071 (-.192, .387)	.276 (.262, .399)	.073 (.056, .245)	.954 (.824, .980)
		$\boldsymbol{\tau}$ (Min, Max)	-.029 (-.425, .333)	.285 (.277, .337)	.077 (.057, .250)	.956 (.762, .986)

similar results, that is, the values of Bias and MSE for ability and speediness parameters decrease as the number of items increases from $J = 20$ to $J = 40$. Third, based on the results of CP values, almost all parameters are around the nominal level 0.95, and additionally, most SD and SE values are comparable to each other, which shows that our MCMC sampling can obtain a very good recovery of the truth for each parameter.

5.2. Performance of the Proposed Criteria

Since our MCMC algorithm yields the satisfactory results of Bayesian estimation for our proposed joint model, next we investigate the empirical performance on the proposed decomposition of DIC and LPML. Following Subsection 5.1, we also consider the same four scenarios, i.e., $N = 125, J = 20$; $N = 125, J = 40$; $N = 250, J = 20$ and $N = 250, J = 40$ and run 500 replications for each scenario. Figure 5.1 shows the boxplots of the $\Delta\mathrm{DIC}_{\mathrm{AC}}$, $\Delta\mathrm{DIC}^{*}_{\mathrm{AC}}$, $\Delta\mathrm{DIC}_{\mathrm{RT}}$, $\Delta\mathrm{DIC}^{*}_{\mathrm{RT}}$, $\Delta\mathrm{DIC}_{\mathrm{PDE}}$, and $\Delta\mathrm{LPML}_{\mathrm{AC}}$, $\Delta\mathrm{LPML}_{\mathrm{RT}}$, $\Delta\mathrm{LPML}_{\mathrm{PDE}}$ for the four scenarios. From these plots, we see that the values of these assessment criteria become larger when the numbers of items or participants increase. These results are intuitively appealing since there is more information in the data when more items and participants are added. In Figure 5.1, all those values are far away from zero which indicates the criteria support that there are gains in fitting the third part of the data with additional information from the other two parts of the data. The median values of $\Delta\text{DIC}_{\text{AC}}$ ($\Delta\text{DIC}^*_{\text{AC}}$) for 500 replications are 121.565 (16.383), 143.657 (19.285), 246.573 (22.528), and 294.548 (26.431) for $N = 125, J = 20$; $N = 125, J = 40$; $N = 250, J = 20$ and $N = 250, J = 40$, respectively, while the median values of $\Delta\text{LPML}_{\text{AC}}$ for 500 replications are 62.894, 74.075, 125.449, and 149.494, respectively. The median values of $\Delta\text{DIC}_{\text{RT}}$ ($\Delta\text{DIC}^*_{\text{RT}}$) for 500 replications are 37.809 (14.964), 40.919 (18.211), 65.126 (20.210), and 71.151 (25.022), respectively, while the median values of $\Delta\text{LPML}_{\text{RT}}$ are 20.261, 21.930, 33.939, and 37.070, respectively. By comparing $\Delta\text{DIC}_{\text{AC}}$ with $\Delta\text{DIC}^*_{\text{AC}}$ (or $\Delta\text{DIC}_{\text{RT}}$ with $\Delta\text{DIC}^*_{\text{RT}}$), we could see in the simulation data, besides the RT data (or the item response data), the PDE data indeed helps a lot in the fit for the joint model. Moreover, the median values of $\Delta\text{DIC}_{\text{PDE}}$ are 144.630, 166.949, 290.020, and 336.068, respectively, while the median values of $\Delta\text{LPML}_{\text{PDE}}$ are 74.089, 84.335, 147.005, and 169.241, respectively.

Further, we have examined the empirical performance of the proposed criteria under three different parameter settings in equation 2.9: (i) $\beta_1 = 0$ and $\beta_2 = 0$ (i.e., the PDE model is independent of the AC model and the RT model), (ii) $\beta_1 = 0$ and $\beta_2 = 3.5$, and (iii) $\beta_1 = 14.5$ and $\beta_2 = 0$. For these three settings, we generate the simulation data using $N = 250$ and $J = 20$ and the true values of all other parameters are set to be the same as those in Section 5.1. The results of the different DICs and LPMLs are reported in Table 5.3 for the three settings of (β_1, β_2). We see from Table 5.3 that under the setting (i), the respective ranges of $\Delta\text{DIC}_{\text{PDE}}$, $\Delta\text{LPML}_{\text{PDE}}$, $\Delta\text{DIC}_{\text{AC}} - \Delta\text{DIC}^*_{\text{AC}}$, and $\Delta\text{DIC}_{\text{RT}} - \Delta\text{DIC}^*_{\text{RT}}$ include zero which is expected since there is no relationship between the PDE model and the rest two model components in the joint model. Further, under the setting (ii), from equation 2.9, we see that the PDE data should help improve the fit of the RT data but not the AC data, which are precisely confirmed empirically by the results shown in Table 5.3, since the range of $\Delta\text{DIC}_{\text{AC}} - \Delta\text{DIC}^*_{\text{AC}}$ includes zero while the range of $\Delta\text{DIC}_{\text{RT}} - \Delta\text{DIC}^*_{\text{RT}}$

Table 5.3: The Range (Min, Max) Results of Model Assessment Criteria

	(i) $\beta_1 = 0, \beta_2 = 0$	(ii) $\beta_1 = 0, \beta_2 = 3.5$	(iii) $\beta_1 = 14.5, \beta_2 = 0$
$\Delta\mathrm{DIC}_{\mathrm{AC}}$	(5.196, 39.722)	(6.685, 50.640)	(213.487, 341.163)
$\Delta\mathrm{DIC}^*_{\mathrm{AC}}$	(6.734, 40.051)	(7.079, 40.923)	(7.387, 41.957)
$\Delta\mathrm{DIC}_{\mathrm{AC}} - \Delta\mathrm{DIC}^*_{\mathrm{AC}}$	(-2.037, 11.292)	(-1.968, 28.621)	(196.374, 320.359)
$\Delta\mathrm{LPML}_{\mathrm{AC}}$	(4.673, 21.936)	(5.417, 27.693)	(108.792, 172.787)
$\Delta\mathrm{DIC}_{\mathrm{RT}}$	(3.924, 39.271)	(25.570,106.766)	(6.428, 38.282)
$\Delta\mathrm{DIC}^*_{\mathrm{RT}}$	(5.868, 38.937)	(5.840, 38.701)	(5.885, 39.240)
$\Delta\mathrm{DIC}_{\mathrm{RT}} - \Delta\mathrm{DIC}^*_{\mathrm{RT}}$	(-2.005, 8.209)	(10.404, 81.322)	(-3.034, 15.362)
$\Delta\mathrm{LPML}_{\mathrm{RT}}$	(3.661, 20.706)	(14.747, 55.110)	(4.313, 20.821)
$\Delta\mathrm{DIC}_{\mathrm{PDE}}$	(-4.146, 9.319)	(8.404, 80.754)	(219.660, 340.188)
$\Delta\mathrm{LPML}_{\mathrm{PDE}}$	(-2.266, 4.862)	(4.186, 40.416)	(111.297, 172.808)

does not include zero. Moreover, under the setting (iii), the PDE data should help improve the fit of the AC data but not the RT data, which is consistent with the results given in Table 5.3 since the range of $\Delta\mathrm{DIC}_{\mathrm{AC}} - \Delta\mathrm{DIC}^*_{\mathrm{AC}}$ does not include zero while the range of $\Delta\mathrm{DIC}_{\mathrm{RT}} - \Delta\mathrm{DIC}^*_{\mathrm{RT}}$ includes zero. In addition, we note that under all of the three simulation settings, we set $\varphi = 0.411$, implying that the AC data are associated with the RT data, and hence, the AC data should help improve the fit of the RT data and vice versa. We see from Table 5.3 that the lower bounds of the ranges of $\Delta\mathrm{DIC}_{\mathrm{AC}}$, $\Delta\mathrm{DIC}^*_{\mathrm{AC}}$, $\Delta\mathrm{DIC}_{\mathrm{RT}}$, and $\Delta\mathrm{DIC}^*_{\mathrm{RT}}$ are larger than zero as they should be. Finally, from Table 5.3, we can see that when the absolute value of β_1 (or β_2) gets bigger, the value of $\Delta\mathrm{DIC}_{\mathrm{PDE}}$ and $\Delta\mathrm{LPML}_{\mathrm{PDE}}$) increase. These results further demonstrate good performance of our proposed model assessment criteria.

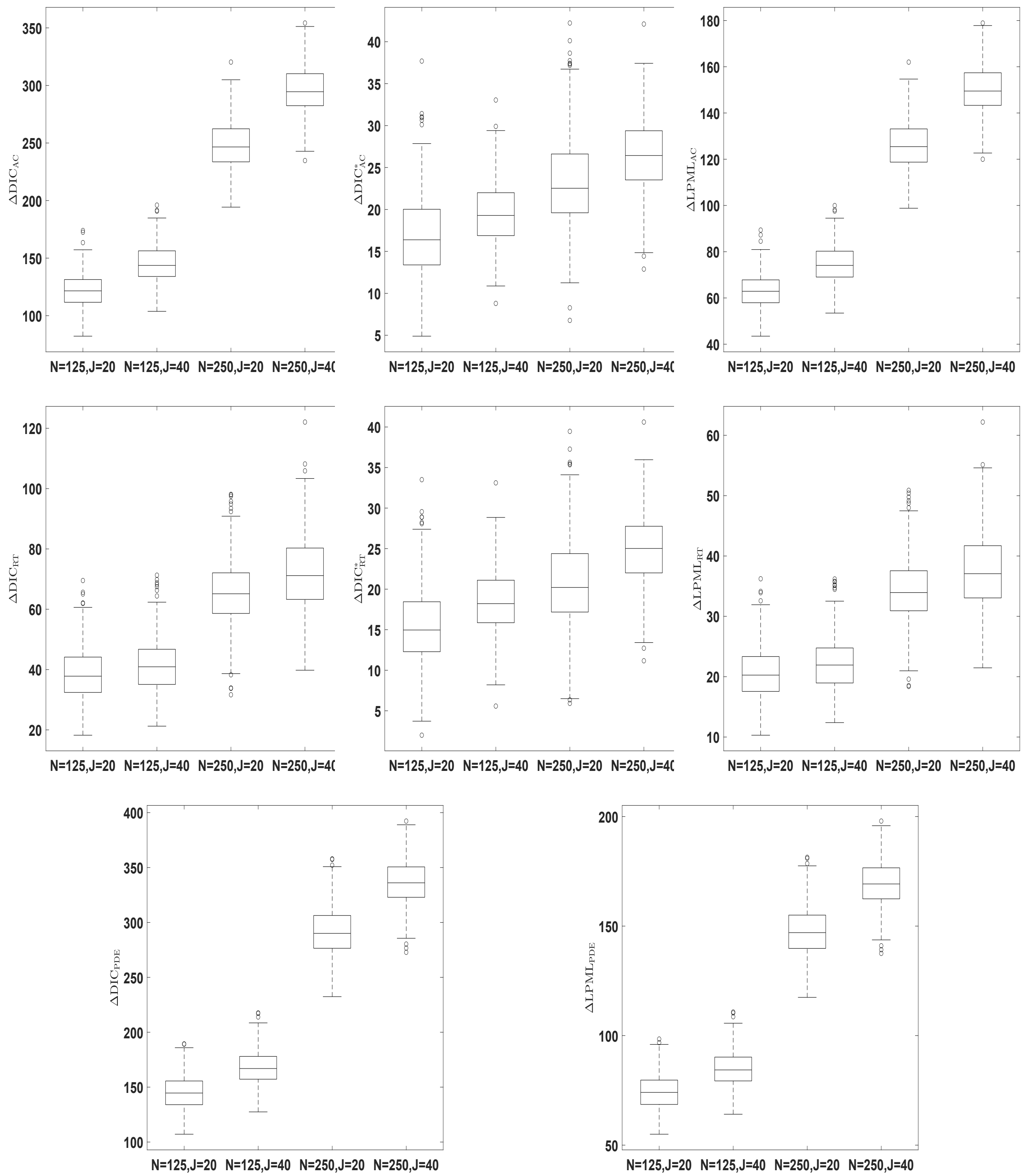

Figure 5.1: $\Delta \mathrm{DIC}_{\mathrm{AC}}$, $\Delta \mathrm{DIC}^*_{\mathrm{AC}}$, $\Delta \mathrm{LPML}_{\mathrm{AC}}$, $\Delta \mathrm{DIC}_{\mathrm{RT}}$, $\Delta \mathrm{DIC}^*_{\mathrm{RT}}$, $\Delta \mathrm{LPML}_{\mathrm{RT}}$, $\Delta \mathrm{DIC}_{\mathrm{PDE}}$, and $\Delta \mathrm{LPML}_{\mathrm{PDE}}$ results for different samples sizes of items and respondents.

6. Empirical Analysis

The AppRISE data consists of the item response and the RT for 43 items from 117 participants, and the PDE score from these participants. There are 65 participants at 5 years old, 51 participants at 6 years old and 1 participant at 7 years old. Among all participants, 51 are female, while 66 are male. The distributions of item accuracies, response times, and PDE scores are shown in Figure S.4 (a)-(c) of the Supplementary Materials, respectively.

In our study, there are two major goals of analyzing the AppRISE and PDE test data. The first major goal is to investigate whether any two parts of this multidimensional data can contribute to the fit of the remaining part of the data. Also, we specifically investigate whether the AC model can be improved by only including the RT data. The second major goal is to establish criterion validity of the tablet test assessment, based on the correlation between PDE scores and ability.

6.1. Bayesian Model Assessment

To compare Bayesian model assessment criteria, we analyze the multidimensional data from AppRISE and PDE tests in four situations: (1) apply the AC model in equation 2.7 alone to the item response data; (2) apply the RT model in equation 2.2 alone to the RT data; (3) apply the model $\mathcal{N}(\mu_{\text{PDE}}, \sigma^2_{\text{PDE}})$ alone to the paper-based PDE data; and (4) apply the joint model from equation 2.7 to equation 2.9 to the item response, RT and PDE data together. In Table 6.1, we present the total values and decomposition of DIC, p_D, and LPML as well as the results of $\Delta\text{DIC}_{\text{AC}}$, $\Delta\text{DIC}^*_{\text{AC}}$, $\Delta\text{DIC}_{\text{RT}}$, $\Delta\text{DIC}^*_{\text{RT}}$, $\Delta\text{DIC}_{\text{PDE}}$, $\Delta\text{LPML}_{\text{AC}}$, $\Delta\text{LPML}_{\text{RT}}$, and $\Delta\text{LPML}_{\text{PDE}}$. It is clearly seen that all values of $\Delta\text{DIC}_{\text{AC}}$, $\Delta\text{DIC}_{\text{RT}}$, $\Delta\text{DIC}_{\text{PDE}}$, $\Delta\text{LPML}_{\text{AC}}$, $\Delta\text{LPML}_{\text{RT}}$, and $\Delta\text{LPML}_{\text{PDE}}$ are positive, which suggests the joint model is supported by the data and shows the additional data (i.e., the availability of the RT and PDE data, the RT and item response data, the item response and PDE data) provides much information in fitting the third part of the data. Those results indicate to incorporate the RT and PDE data into the joint model might be an important source for refining our estimations of the AC model. In addition, the positive value for $\Delta\text{DIC}^*_{\text{AC}}$ shows the introduction of the RT model alone to conjointly modeling the item response data indeed provides the gain in the fit of the AC model. Similarly, by incorporating the item response data alone in the joint modeling with the RT data, it truly helps the fit for the RT model, which is verified through the positive value of $\Delta\text{DIC}^*_{\text{RT}}$.

6.2. Posterior Estimation

To analyze the multidimensional data from AppRISE tablet assessment and PDE paper-and-pencil test, we use the same prior specification as in Section 5 for all unknown parameters in the model. For Bayesian analysis, we have run $60,000$ MCMC samples with a burn-in period of $10,000$ iterations for all four situations and then we have thinned the MCMC samples for every 5 steps to compute the posterior estimates. We have looked at the traceplots and autocorrelated plots of all unknown parameters for informally checking about the convergence and found all parameters are converged after 10,000 iterations. We found the posterior means of b_{13} and b_{42} (2.106 and 2.151, respectively) have the largest values compared to other items, which is expected as these two items have the lowest correct rates among all items. The posterior estimates of a_js, b_js , λ_js, θ_is and τ_is are shown in Table S.2 to Table S.8 of the Supplementary Materials.

To examine whether the paper-based PDE test and the tablet-based AppRISE test assess the similar latent construct, we have calculated the empirical correlation between PDE_is and θ^*_is, which has a posterior mean of 0.867 with a 95% HPD interval $(0.773, 0.941)$. This suggests the PDE are highly correlated with the latent ability θ^*_is used in the AC model for the AppRISE data. However, the empirical correlation σ_{12} between θ^*_is and τ^*_is has a posterior mean of 0.134 with a 95% HPD interval $(0.062, 0.199)$, which implies the correlation between the latent ability and speediness of a participant is

Table 6.1: The Results of Model Assessment Criteria

Situations	DIC	p_D	LPML
(1) $\mathrm{AC}_{\mathrm{only}}$	2496.413	45.119	-1254.638
(2) $\mathrm{RT}_{\mathrm{only}}$	4493.838	79.809	-2267.229
(3) $\mathrm{PDE}_{\mathrm{only}}$	1006.761	1.942	-503.521
(4) Joint Model	7854.472	132.911	-3954.526
$[\mathrm{AC} \mid \mathrm{RT}, \mathrm{PDE}]$	2390.392	49.168	-1199.618
$[\mathrm{RT}, \mathrm{PDE}]$	5464.080	83.743	-2754.908
$[\mathrm{RT} \mid \mathrm{AC}, \mathrm{PDE}]$	4458.223	82.887	-2247.706
$[\mathrm{AC}, \mathrm{PDE}]$	3396.249	50.024	-1706.820
$[\mathrm{PDE} \mid \mathrm{AC}, \mathrm{RT}]$	873.551	3.463	-434.834
$[\mathrm{AC}, \mathrm{RT}]$	6980.921	129.448	-3519.692
$[\mathrm{AC} \mid \mathrm{RT}]$	2484.589	48.434	-
$[\mathrm{RT}]$	4496.332	81.014	-
$[\mathrm{RT} \mid \mathrm{AC}]$	4481.462	81.914	-
$[\mathrm{AC}]$	2499.459	47.534	-
$\Delta\mathrm{DIC}_{\mathrm{AC}}$	$\Delta\mathrm{DIC}^*_{\mathrm{AC}}$	$\Delta\mathrm{LPML}_{\mathrm{AC}}$	$\Delta\mathrm{DIC}_{\mathrm{PDE}}$
106.021	11.824	55.020	133.210
$\Delta\mathrm{DIC}_{\mathrm{RT}}$	$\Delta\mathrm{DIC}^*_{\mathrm{RT}}$	$\Delta\mathrm{LPML}_{\mathrm{RT}}$	$\Delta\mathrm{LPML}_{\mathrm{PDE}}$
35.615	12.376	19.523	68.687

Table 6.2: The posterior estimations of parameters for modeling the PDE scores

Parameter	Posterior Mean	HPD Interval	Parameter	Posterior Mean	HPD Interval
β_0	93.820	$[90.826, 96.752]$	β_1	14.563	$[12.052, 17.065]$
β_2	3.534	$[1.122, 5.959]$	σ^2_{PDE}	57.188	$[24.882, 89.070]$

relatively low in this test. In addition, σ_{23}, which is the covariance between τ_i^* and ζ_i^*, has a posterior mean of 3.056 with a 95% HPD interval $(1.889, 4.220)$. While for the variance of ζ_i^*, σ_{33} has a posterior mean of 284.867 with a 95% HPD interval $(222.009, 353.544)$. Notice that to present the results, we have transformed θ_is and τ_is used in the computation back to their original definitions θ_i^*s and τ_i^*s in the joint model. We can further justify a common latent ability for the PDE test and the AppRISE test by establishing a hypothesis test that $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, as $\beta_1 = 0$ indicates that the construct assessed by two tests have no correlations. By Lindley's method (Lindley, 1965), we can test this hypothesis in an ad hoc way and we reject the hypothesis of $\beta_1 = 0$ at the significance level $\alpha = 5\%$ since the 95% HPD interval of β_1 does not include zero (c.f., Table 6.2). This provides additional evidence to support the PDE test and AppRISE test can be used to assess a common latent ability. In Table 6.2, the 95% HPD interval for the coefficient β_2 in the PDE normal model is also far away from zero with a posterior mean estimate of 3.534, which indicates the speediness of a participant also have an impact on the traditional paper-based test.

In Section S.2 of the Supplementary Materials, we propose a Bayesian procedure to compare and order item difficulties, which will help rank the item difficulty in the AppRISE. The result for ordering the item difficulty can be used to improve the item bank and the adaptive testing design of the AppRISE tablet assessment in the future. In Section S.3, we develop Bayesian residuals for assessing the model adequacy of the joint model in fitting the item response, RT and PDE data, which shows the distribution assumption of the proposed joint model in equation 2.1 to equation 2.4 (or equivalent to equation 2.7 and equation 2.9) are reasonable. A calibration algorithm for quantifying uncertainty of the DIC and LPML assessment criteria is developed in Section S.5. The details of these developments and the analysis results of the data from the AppRISE tablet test and PDE paper-based test are given in the Supplementary Materials.

7. Conclusion and Discussion

In this paper, a joint model is proposed for the item response and the response time from the AppRISE tablet assessment, and the age-normed standard score of Phonemic Decoding Efficiency (PDE) from a paper-and-pencil test. The reformulation and reparameterization of the proposed joint model is one of the major developments in this paper as it facilitates more convenient specification of the prior distributions and makes it possible to implement MCMC sampling from the joint posterior distribution using an R package called *nimble*. Another major development in this paper is the decomposition of DIC and LPML, which enable us to separately evaluate the contribution of different sources of the data in a joint model. Also, we have put forward novel Bayesian criteria, $\Delta\text{DIC}_{\text{AC}}$, $\Delta\text{DIC}^*_{\text{AC}}$, $\Delta\text{DIC}_{\text{RT}}$, $\Delta\text{DIC}^*_{\text{RT}}$, $\Delta\text{DIC}_{\text{PDE}}$, $\Delta\text{LPML}_{\text{AC}}$, $\Delta\text{LPML}_{\text{RT}}$, and $\Delta\text{LPML}_{\text{PDE}}$, which can be used to assess the gain of modeling one part data by incorporating additional parts data. Both the simulation and the real data analysis demonstrated that those novel criteria are effective and promising. Moreover, we have proposed some new ideas to rank item difficulty for AC models with uncertainty and to assess the model adequacy for the RT, item response and PDE data.

In addition, as demonstrated in de la Torre and Patz (2005), in the IRT model with a multidimensional ability, the benefit of correlated dimensions is that the correlations among the dimensions provide additional information, which may lead to more precise ability estimates for each dimension. Based on our proposed approach, this similar type of benefits for correlated dimensions can be assessed by using our difference measures such as $\Delta\text{DIC}_{\text{AC,PDE}}$, $\Delta\text{DIC}_{\text{AC,PDE}}$, and $\Delta\text{DIC}_{\text{RT,PDE}}$ for the AppRISE data. These difference measures can quantify the gain in the fit of certain correlated multidimensional data by incorporating the data from an additional correlated dimension. Table S.10 of the Supplementary Materials show the values of the difference measures for the two dimensional data by incorporating the data from the third dimension for the AppRISE data, which are quite consistent with the posterior estimates of the covariance and correlation parameters for all of the three correlated dimensions (AC, RT, PDE) given in Table S.9 of the Supplementary Materials. These results provide a further justification of modeling the item responses, RTs, and PDE scores jointly.

Finally, the decomposition of DIC can be extended to the decomposition of DIC_L proposed by Li et al. (2020). Recently, WAIC (Watanabe, 2010) has become a popular Bayesian model comparison criterion. WAIC is constructed based on the posterior predictive density, which bears a resemblance with LPML. Thus, the decomposition of LPML can be extended to the decomposition of WAIC. However, these two extensions are quite extensive in terms of both analytical derivations and computational developments, which is beyond the scope of the current paper. Therefore, these extensions are deserved to be another interesting topic of future research.

Acknowledgement

We would like to thank the Executive Editor, the Associate Editor, and three reviewers on their valuable suggestions and comments, which have led to a much improved version of the paper. The work was done when F. Liu visited the University of Connecticut as a visiting scholar. This research was partially supported by the University of California Office of the President and the University of Connecticut InCHIP seed grant. Dr. Wang's research was also partially supported by the US National Science Foundation under Grant No. 1848451.

References

Bolsinova, M., de Boeck, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika*, *82*(4), 1126–1148.

Bolt, D. M., Wollack, J. A., & Suh, Y. (2012). Application of a multidimensional nested logit model to multiple-choice test items. *Psychometrika*, *77*(2), 339–357.

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*(4), 651–673.

Chan, J. C., & Grant, A. L. (2016). Fast computation of the deviance information criterion for latent variable models. *Computational Statistics & Data Analysis*, *100*, 847–859.

Chen, G., & Luo, S. (2018). Bayesian hierarchical joint modeling using skew-normal/independent distributions. *Communications in Statistics-Simulation and Computation*, *47*(5), 1420–1438.

Chen, M.-H., & Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, *8*(1), 69–92.

Chen, M.-H., Shao, Q.-M., & Ibrahim, J. G. (2000). *Monte carlo methods in Bayesian computation*. Springer Science & Business Media.

de la Torre, J., & Patz, R. J. (2005). Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics*, *30*(3), 295–311.

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., ... Paganin, S. (2020). *NIMBLE: MCMC, particle filtering, and programmable hierarchical modeling.* Retrieved from `https://cran.r-project.org/package=nimble` doi: 10.5281/zenodo.1211190

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, *26*(2), 403-413.

Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods*, *41*(4), 1095–1110.

Entink, R. K., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, *74*(1), 21-48.

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

Fox, J.-P., & Marianti, S. (2016). Joint modeling of ability and differential speed using responses and response times. *Multivariate Behavioral Research*, *51*(4), 540–553.

Fujimoto, K. A. (2018). A general Bayesian multilevel multidimensional IRT model for locally dependent data. *British Journal of Mathematical and Statistical Psychology*, *71*(3), 536–560.

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

Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B*, *56*(3), 501–514.

Gelfand, A. E., Dey, D. K., & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based-methods (with discussion). In A. P. D. J.M. Bernado J.O. Berger & A. Smith (Eds.), *In bayesian statistics 4.* Oxford: Oxford University Press.

Gilbert, J. K., Compton, D. L., Fuchs, D., & Fuchs, L. S. (2012). Early screening for risk of reading disabilities: Recommendations for a four-step screening system. *Assessment for Effective Intervention*, *38*(1), 6–14.

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

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika*, *68*(4), 563–583.

Karadavut, T. (2019). The uniform prior for bayesian estimation of ability in item response theory models. *International Journal of Assessment Tools in Education*, *6*(4), 568–579.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Li, Y., Yu, J., & Zeng, T. (2020). Deviance information criterion for latent variable models and misspecified models. *Journal of Econometrics*, *216*(2), 450–493.

Lindley, D. V. (1965). *Introduction to probability and statistics from a bayesian viewpoint.* Cambridge University Press.

Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction time and accuracy in psycholinguistic experiments. *Psychometrika*, *76*(3), 487–503.

Lu, J., Wang, C., Zhang, J., & Tao, J. (2020). A mixture model for responses and response times with a higher-order ability structure to detect rapid guessing behaviour. *British Journal of Mathematical and Statistical Psychology*, *73*(2), 261-288.

Luce, R. (1991). *Response times: Their role in inferring elementary mental organization.* OUP USA.

Man, K., Harring, J. R., Jiao, H., & Zhan, P. (2019). Joint modeling of compensatory multidimensional item responses and response times. *Applied Psychological Measurement*, *43*(8), 639–654.

Merkle, E. C., Furr, D., & Rabe-Hesketh, S. (2019). Bayesian comparison of latent variable models: Conditional versus marginal likelihoods. *Psychometrika*, *84*(3), 802–829.

Molenaar, D., & de Boeck, P. (2018). Response mixture modeling: Accounting for heterogeneity in item characteristics across response times. *Psychometrika*, *83*(2), 279–297.

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*(2), 491–513.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, *64*(4), 583–639.

Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *Test of word reading efficiency:(TOWRE-2).* Pearson Clinical Assessment.

van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, *46*(3), 247-272.

van der Linden, W. J. (2017). *Handbook of item response theory, volume three: applications.* Chapman and Hall/CRC.

van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, *73*(3), 365–384.

van der Linden, W. J., & Hambleton, R. K. (2013). *Handbook of modern item response theory.* Springer Science & Business Media.

Visual Numerics, I. (2003). *Imsl fortran library user's guide math/library.* San Ramon, CA: Visual Numerics, Inc.

Wang, X., Saha, A., & Dey, D. K. (2016). *Bayesian joint modeling of response times with dynamic latent ability in educational testing* (Vol. 3; Tech. Rep.). Storrs, Connecticut: Department of Statistics, University of Connecticut.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

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, F., Chen, M.-H., Cong, X. J., & Chen, Q. (2021). Assessing importance of biomarkers: A bayesian joint modelling approach of longitudinal and survival data with semi-competing risks. *Statistical Modelling*, *21*(1-2), 30–55.

Zhang, X., Tao, J., Wang, C., & Shi, N.-Z. (2019). Bayesian model selection methods for multilevel IRT models: A comparison of five DIC-based indices. *Journal of*

Educational Measurement, *56*(1), 3–27.