Psychometrika (2026), **91**, 1361–1390
doi:10.1017/psy.2026.10128

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Longitudinal Designs for Diagnostic Models: Identification and Estimation

Trung Tran Quoc Le[1], Steven Andrew Culpepper[2] and Jeff Douglas[2]

[1]Department of Psychology, University of Illinois Urbana-Champaign, USA; [2]Department of Statistics, University of Illinois Urbana-Champaign, USA

Corresponding author: Trung Le; Email: trungle467@gmail.com

(Received 1 September 2025 UTC; revised 5 February 2026 UTC; accepted 1 June 2026 UTC)

Abstract

Recent studies on cognitive diagnostic models (CDMs) have extended the framework to longitudinal data. Various methods combined CDMs and hidden Markov models (HMMs) to assess changes in attributes over time and to evaluate the effects of interventions and covariates. A requirement for model fitting, inference, and interpretation is that models are identifiable. In this article, we derive identifiability conditions for experimental design used in education research. Specifically, we consider three designs: pretest/posttest single-group design, counterbalancing, and multiple-group longitudinal design. Drawing on existing HMM research, we examine the extent to which these designs satisfy common identifiability assumptions and propose new constraints for the counterbalancing and multiple-group longitudinal designs. These two setups are recommended in situations where item parameters differ over time or when there is heterogeneity in transition patterns across groups. We introduce a general HMM model for the multiple-group longitudinal design and a Gibbs sampling algorithm to estimate the parameters. We assess parameter recovery through a Monte Carlo simulation study and apply the model to a dataset from a study which evaluates the effectiveness of two interventions relative to a control condition. The results demonstrate the flexibility of the model and its potential to offer new insights into learning processes.

Keywords: cognitive diagnosis model; hidden Markov model; identifiability; longitudinal research design

1. Introduction

There has been an increasing interest in longitudinal data as they are powerful tools to track changes and development. In educational and psychological settings, one of the central topics is measuring individuals' development of latent attributes over time. This is particularly important for assessing the effectiveness and impact of interventions on outcomes. Recently, hidden Markov models (HMMs) are becoming popular in psychology and education research as a method for modeling changes between attributes over time (Chen et al., 2018; Wang et al., 2018). HMMs are widely used in fields such as signal processing (Crouse et al., 2002) and information theory (Simola et al., 2008) because of their ability to draw inferences about important unobservable states. In the context of educational and psychological testing, researchers have adapted the HMM framework to perform cognitive diagnosis using a class of models that will be referred to as restricted HMMs (RHMMs). RHMMs combine features of HMMs with a restricted latent class model (RLCM; Xu, 2017), where the RLCM imposes structure on the emission probability—the conditional probability of a response given current hidden state. RHMMs,

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

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

which integrate RLCM with HMM, along with the conventional HMMs have the potential to advance educational and psychological research, and help identify interventions that can enhance outcomes (Ye et al., 2016).

It is, thus, crucial to examine issues such as the identifiability conditions for these RHMMs to ensure accurate parameter recovery and reliable diagnostic inferences. Several studies have investigated the identifiability of HMMs (Allman et al., 2009; Bonhomme et al., 2016), along with other research focusing on the identifiability of cognitive diagnostic models (CDMs; Xu, 2017; Xu & Shang, 2018) and RHMMs (Chen et al., 2015, 2020; Liu et al., 2023). Methods for assessing model identifiability are applied under various conditions, as the structure of data can differ depending on the experimental designs, leading to different model specifications. Existing identifiability conditions for HMMs and RHMMs, which are derived under specific assumptions, may not be feasible in all settings. The focus of this article is to examine longitudinal settings (HMMs and RHMMs), evaluate existing model assumptions, and derive new identifiability results for different research designs that may be easier to satisfy in practice. Particularly, we examine three experimental designs: single-group pretest/posttest design, counterbalancing, and multiple-group longitudinal design.

The three experimental setups are used to evaluate the efficacy of interventions; however, each design relies on distinct assumptions which result in different modeling approaches. Motivation for choosing the three specified research designs comes from the following studies in the field of educational testing. Wang et al. (2018) analyzed data from a counterbalancing study using a higher-order HMM combined with a constrained CDM. Their model assumes that once a student masters an attribute, they will never lose it. In contrast, Kaya and Leite (2017) and Madison and Bradshaw (2018a), in studies using single-group pretest/posttest design, allowed attribute levels to drop between two time points. The CDMs applied in Kaya and Leite (2017), Madison and Bradshaw (2018a), and Wang et al. (2018) all require correct specification of a Q-matrix, a binary matrix that defines the relationships between the questions (e.g., test items) and the underlying attributes being measured. Similarly, Tang and Zhan (2021) as well as Madison and Bradshaw (2018b) developed models by integrating HMM with constrained CDMs for a study using a multiple-group longitudinal design. Tang and Zhan (2021) allowed the item parameters at the first time point to be freely estimated while constraining those for later time points to ensure the comparability of item parameter estimates across groups and time points. On the other hand, the model proposed by Madison and Bradshaw (2018b) can account for different item parameters between groups.

Recent advancements in existing software packages have allowed easier access for practitioners to estimate the parameters of RHMMs and apply the models (Madison et al., 2025). However, it should be noted that the RHMMs applied by Tang and Zhan (2021), Madison and Bradshaw (2018a, 2018b), Kaya and Leite (2017), and Wang et al. (2018) assume specific structures to the item responses and impose certain parameter constraints across items and attributes. Violations of these assumptions can result in inaccurate parameter estimates and misclassification of individual's latent profiles. While these studies have significantly contributed to the development of methodologies for longitudinal designs, their focus was not on investigating model identifiability. There has been considerable research on identifiability of HMMs; however, few studies have extended these findings to RHMMs as used in the experimental designs discussed above. For instance, Liu et al. (2023) provide identifiability constraints for RHMMs with general RLCM and focus on situations where emission probabilities are formed using a deterministic inputs, noisy "and" gate (DINA; Junker & Sijtsma, 2001). The conditions derived by Liu et al. (2023) may apply to the single-group pretest/postest design, but may not extend to the other two experimental setups. Accordingly, the current work builds on existing research in identifying HMMs and RHMMs by specifying sufficient conditions for the remaining two designs: counterbalancing and multiple-group longitudinal designs.

In this article, we investigate HMMs with discrete time and finite observables and hidden states. Conventional HMMs consist of two components: 1) an emission matrix, **B**, that describes the probability of observed responses given latent states and 2) a transition matrix, **A**, that denotes the likelihood of changing from one state to another over two adjacent time points. In existing identifiability research, **B** and **A** are typically assumed to be constant over time along with an absence of an absorbing

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

state. These requirements might not be satisfied depending on the research design employed and the associated assumptions. The two experimental setups examined in our article are situations where these specific assumptions might not be practical. Therefore, we want to relax these assumptions and impose other restrictions by considering HMMs with time-varying emission matrices as well as group-specific transition matrices. Based on the experimental designs, we provide new identification proofs for HMMs and RHMMs when **B** changes over time and **A** differs between groups. The identifiability for HMMs discussed in our article is defined up to a permutation on the latent state labels. From the novel identifiability theorems, we demonstrate a Bayesian formulation and present a Gibbs sampling algorithm to estimate parameters for generalized HMMs in the case of the multiple-group longitudinal design. We demonstrate the flexibility and potential of our algorithms as well as the conventional HMMs through a data application designed to study the effectiveness of a learning intervention.

The remainder of the article is organized as follows. We first discuss existing HMM research in education and psychology in addition to introducing the typical setup of conventional HMMs and RHMMs. We then examine the counterbalancing and multiple-group longitudinal designs and generalize the results in Bonhomme et al. (2016) to establish identifiability constraints. Next, we discuss how the new identifiability conditions may inform practitioners to design studies under HMM/RHMM framework. Afterward, we present a Bayesian formulation of generalized HMMs for the multiple-group longitudinal design and propose a Gibbs sampling algorithm for posterior approximation. The fifth section reports results of a Monte Carlo simulation study which illustrate the accuracy of the proposed algorithm. The algorithm is then applied to the dataset in Tang and Zhan (2021) to demonstrate its ability to assess transitions between attributes in the evaluation of two interventions with respect to a control condition. The last section summarizes the findings of the present article and provides concluding remarks. All proofs and technical details are provided in the appendix.

2. Identifiability for HMM

This section examines identifiability conditions for three research designs: single-group pretest/posttest design, counterbalancing, and multiple-group longitudinal design. In a single-group pretest/posttest design, individuals complete evaluations before and after instruction, with both assessments typically using similar items. In contrast, counterbalancing and multiple-group longitudinal designs may involve more than two groups and varying items over time. We consider multiple-group longitudinal design to be experiments in which individuals are assigned to treatment and control groups. Performances are compared across groups, with the control group serving as the baseline, where minimal intervention occurs. Counterbalancing, on the other hand, involves administering a predetermined set of questions in different orders. The orders correspond to different test versions and participants are assigned to groups based on the test version they receive.

We begin the discussion on identifiability by presenting an overview of existing HMM research in education and psychology research. The first section also introduces the notation for conventional HMMs and the identifiability results from Liu et al. (2023) for the single-group pretest/postest design. Subsequent sections extend the definitions to the counterbalancing and multiple-group longitudinal designs, providing identifiability conditions for each. The last section provides comments and remarks on the general approach used to derive these conditions.

2.1. Overview of existing HMM research in education and psychology

Consider a general HMM model for an observation at time points $t = 1, \ldots, T$. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)^\top$ denote the response data of an individual over time. Let $[q]$ for $q \in \mathbb{N}$ denote the set $\{1, 2, \ldots, q\}$. Here, $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{Jt})$ is the response pattern of an individual for J items at time t and $\boldsymbol{X}_t \in ([M_1] \times [M_2] \times \cdots \times [M_J])$. In other words, M_j is the number of response options for an item j. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_T)^\top$ be the latent states of an individual over time and $Z_t \in [r]$ for $r \in \mathbb{N}$. The response patterns $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T$ are

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

assumed to be conditionally independent given $Z_1, \ldots, Z_T$. In other words, each $\boldsymbol{X}_t$ is directly impacted by the corresponding Z_t and the conditional distribution is

$$P(\boldsymbol{X}_t \mid Z_1, \ldots, Z_t) = P(\boldsymbol{X}_t \mid Z_t).$$

Furthermore, within a first-order HMM, the distribution of Z_t depends only on the previous state of the immediate past Z_{t-1}. That is,

$$P(Z_t \mid Z_1, \ldots, Z_{t-1}) = P(Z_t \mid Z_{t-1}).$$

HMMs consist of two components. The first element is a transition matrix that dictates the likelihood of changing between latent states over two adjacent time points. Let $\mathbf{A}$ be an $r \times r$ transition matrix with each element (z_{t-1}, z_t) represents

$$a_{z_{t-1}, z_t} = P(Z_t = z_t \mid Z_{t-1} = z_{t-1}),$$

which is the probability of transition to state $Z_t = z_t$ given the membership during the previous time is $Z_{t-1} = z_{t-1}$. Observe that each element in $\mathbf{A}$ is non-negative and each row sums to one.

Additionally, there is an emission matrix that describes the distribution of observed responses given contemporaneous latent states. Let $\mathbf{B}$ be the $q \times r$ emission matrix with each element $(\boldsymbol{x}_t, z_t)$ denoted by

$$b_{\boldsymbol{x}_t, z_t} = P(\boldsymbol{X}_t = \boldsymbol{x}_t \mid Z_t = z_t).$$

That is, each element in $\mathbf{B}$ represents the conditional probability of a response pattern given contemporaneous hidden state $Z_t = z_t$. The elements of $\mathbf{B}$ are also non-negative and the columns sum to one.

We define $\boldsymbol{\pi}_t = (\pi_{t,1}, \ldots, \pi_{t,r})$, where π_{t,z_t} is the probability of being in state $Z_t = z_t$ at time t. Then, $\boldsymbol{\pi}_1$ is the initial distribution of the latent states. Observe that $\sum_{z_t=1}^{r} \pi_{t,z_t} = 1$ for any $t \in [T]$. Using the initial distribution, we can iterate forward and find the distribution for time $t = 2$, $\boldsymbol{\pi}_2$, to have elements

$$P(Z_2 = z_2) = \sum_{z_1=1}^{r} P(Z_1 = z_1) P(Z_2 = z_2 \mid Z_1 = z_1).$$

We can write $\boldsymbol{\pi}_2 = \boldsymbol{\pi}_1 \mathbf{A}$. Consequently, iterating forward to time n implies that the distribution of latent states is $\boldsymbol{\pi}_n = \boldsymbol{\pi}_1 \mathbf{A}^{n-1}$. The stationary (or long-run) distribution, which characterizes the marginal probability of residing in state z_t, is defined as $\boldsymbol{\pi} = \lim_{n \to \infty} \boldsymbol{\pi}_1 \mathbf{A}^n = (\pi_1, \ldots, \pi_r)$.

For HMMs and RHMMs, a model is identifiable when different values of the parameter set $(\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}_1)$ correspond to distinct marginal distributions of the observed variables. Model identifiability is essential for ensuring that the underlying parameters can be consistently estimated, thereby enabling the model to accurately assess the effects of interventions. Allman et al. (2009) provide a generic identifiability condition for HMMs using results in Kruskal (1977). More recently, Bonhomme et al. (2016) establish sufficient conditions for identifying HMMs, which are more feasible to verify in practice.

Theorem 1 (Bonhomme et al., 2016). *Assume* $T \geq 3$. *The matrices* $\mathbf{B}$, $\mathbf{A}$, *and* $\mathbf{D}_{\boldsymbol{\pi}}$ *are all identified if rank* $(\mathbf{B}) = r$ *and rank* $(\mathbf{A}) = r$, *and* $\pi_r > 0$ *for all* r

Here, $\mathbf{D}_{\boldsymbol{\pi}} = \mathrm{diag}(\pi_1, \ldots, \pi_r)$. The proof of Theorem 1 involves rewriting HMMs into mixture models represented as three-way arrays, where the conditional distribution of the responses can be expressed in terms of $\mathbf{B}$, $\mathbf{A}$, and $\boldsymbol{\pi}$. Then, Bonhomme et al. (2016) derived conditions for finite HMMs to be identifiable using results from the simultaneous diagonalization problem (De Leeuw & Pruzansky, 1978). It is worth noting that Theorem 1 applies to general HMMs and requires $\mathbf{B}$ and $\mathbf{A}$ to be full column rank. Other restrictions may be imposed to identify the parameters of an HMM. Liu et al. (2023) and Chen et al. (2020) extend the results to prove identifiability conditions for HMMs and RHMMs by relaxing some assumptions in Bonhomme et al. (2016). Note that Liu et al. (2023) and Chen et al. (2020) focus on HMMs and RHMMs with binary latent attributes. Particularly, let K be the number of binary latent attributes, then the number of latent states is $r = 2^K$ corresponding to all possible attribute profiles.

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Theorem 2 (Liu et al., 2023). *Let K be the total number of latent attributes. Consider the bipartition of all J items into two disjoint, nonempty subsets $\mathbb{J}_1 = [S]$ and $\mathbb{J}_2 = \{S+1, \ldots, J\}$. Then, rewrite $\boldsymbol{X}_t = (\boldsymbol{X}_t^{\mathbb{J}_1\top}, \boldsymbol{X}_t^{\mathbb{J}_2\top})$ for $t \in [T]$. Let $\mathbf{B}^{\mathbb{J}_1}$ and $\mathbf{B}^{\mathbb{J}_2}$ be the emission matrices for $\boldsymbol{X}_t^{\mathbb{J}_1}, \boldsymbol{X}_t^{\mathbb{J}_2}$. We have $\mathbf{B} = \mathbf{B}^{\mathbb{J}_1} * \mathbf{B}^{\mathbb{J}_2}$, where $*$ is the Khatri–Rao product defined in Definition A.2 in Appendix A. The parameters are identifiable if rank $(\mathbf{A}) = 2^K$, $\pi_l > 0$ for all l, and*

(a) *for $T \geq 3$, rank $(\mathbf{B}) = 2^K$;*
(b) *for $T = 2$, rank $(\mathbf{B}^{\mathbb{J}_1}) = 2^K$ and rank$_K(\mathbf{B}^{\mathbb{J}_2}) \geq 2$.*

Note rank $_K$ in part (b) of Theorem 2 denotes the Kruskal rank. The Kruskal rank of a matrix with r columns equals r if and only if the matrix has full column rank (see Definition A.1 in Appendix A for more details). Moreover, observe that certain assumptions are shared between Theorems 1 and 2. Namely, **B** as well as **A** need to be constant over time, and that every element of the stationary distribution is positive. We now examine what these assumptions entail in real-world settings as well as within the model.

2.2. Challenges with current HMMs

Bonhomme et al. (2016), Chen et al. (2020), and Liu et al. (2023) assume **B** and **A** to be time invariant along with every element in $\boldsymbol{\pi}$ to be positive. It is essential to understand the implications of these assumptions. Having constant **B** implies that the conditional distribution $P(\boldsymbol{X}_t \mid Z_t)$ does not change with time. In other words, parallel measurements are observed across different time points. This assumption could be reasonable when a common set of items is used over time, such as diagnosing mental health conditions with a standardized questionnaire. However, the constant **B** assumption is not always feasible, such as when there may be item parameter drift or when different items are administered over time.

The assumption of constant **A** suggests that $P(Z_t \mid Z_{t-1})$ remains the same for all t. That is, the process being modeled is assumed to be stable during the observation period. This assumption is possible when data are collected over a short time span and conditions are expected to remain consistent. However, the constant **A** assumption may be violated in studies involving extended periods or heterogeneous populations, where individual response to interventions may vary. For example, in educational settings, learning interventions may only be effective for certain subgroups, resulting in different transition dynamics between the groups.

The last assumption requires $\boldsymbol{\pi} > 0_r$ with 0_r represents an r-vector of zeros and " >" is an element-wise inequality. This pertains to the irreducibility of **A**, which is violated if **A** contains absorbing states. A state z_{t-1} is an absorbing state if it has a 100% of transitioning to itself and 0% to any other states. Intuitively, an absorbing state is a state such that once entered, cannot be left. The irreducibility assumption can be violated in educational research aimed at learning advancement. In such cases, interventions are designed to enhance skill development toward a mastery state, which may function as an absorbing state if skill forgetting is considered unlikely.

For a single-group pretest/posttest design, the assumptions of constant **B** and **A** are often reasonable as it is common to use similar assessments before and after an intervention. Consider a study in mathematics education examining the effects of the Enhanced Anchored Instruction program (Bottge et al., 2003) on mathematics achievement of middle school students (Li et al., 2016). The study investigates the impact of the EAI intervention by observing student' performance on a test called fraction of the cost (FOC). The same FOC test is administered before and after the intervention. Hence, we can see that the constant **B** and **A** assumptions are satisfied here. Furthermore, the study consists of only two time points, making the likelihood of having a complete mastery state to be unlikely. This means that students can be expected to obtain or retain knowledge as well as forget some knowledge over the short periods. Therefore, the irreducibility assumption is also feasible, allowing Theorem 2 to be applied to this design. However, these assumptions may not hold for situations where item parameters

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Table 1. Balanced block design in Wang et al. (2018)

	Time points				
Groups	1	2	3	4	5
Group 1	Booklet 1	Booklet 2	Booklet 3	Booklet 4	Booklet 5
Group 2	Booklet 2	Booklet 3	Booklet 4	Booklet 5	Booklet 1
Group 3	Booklet 3	Booklet 4	Booklet 5	Booklet 1	Booklet 2
Group 4	Booklet 4	Booklet 5	Booklet 1	Booklet 2	Booklet 3
Group 5	Booklet 5	Booklet 1	Booklet 2	Booklet 3	Booklet 4

change over time and there are multiple groups with heterogeneous transition parameters. The next sections discuss the specific setups and identifiability proofs of two designs that we recommend using in cases where item parameters differ over time or when there are known groups that might differ in the probability of achieving mastery.

2.3. Counterbalancing for longitudinal diagnostic design

There are several advantages for one to consider implementing a counterbalancing design. This design balances out the item positions and controls for potential order effects. Wang et al. (2018) proposed using a counterbalanced booklet design to address confounding factors that could obscure learning effects. Additionally, Wang et al. (2018) noted that counterbalancing can help avoid empirical identifiability issues associated with using RLCMs, specifically the DINA model. Particularly, they provided the example of estimated guessing parameters being inaccurate for items in later time points. This can happen when there is no guessing because all individuals within a sample have mastered the required attributes by the time they get to the items. Nevertheless, the primary aim of Wang et al. (2018) was not to understand the identifiability of HMMs. The current article fills this gap by formally deriving identifiability conditions for RHMMs under a counterbalancing design.

In what we refer to as a counterbalancing design, a fixed number of questions are arranged into different orderings. These orderings correspond to the various versions of the test, and examinees are then categorized into groups depending on the version they receive. That is, counterbalancing here refers to an experimental design in which the same set of items is administered in different orders across examinees to control for potentially confounding factors, such as order and time effects. This design is relatively popular in both classroom and large-scale assessments (Johnson, 1992; Shute et al., 2008). While counterbalancing can be implemented in many ways, we focus only on those configurations that are relevant for ensuring model identifiability, which will be specified later in this article. An example is the experiment conducted by Wang et al. (2018), where participants responded to 50 items organized into five 10-item blocks, with a learning intervention administered between blocks. Here, the five equal-size partitions of the 50 items are referred to as *booklets* 1–5. Counterbalancing was implemented by rotating each of the five booklets to be administered first in the sequence (Table 1).

Following the experimental setup of Wang et al. (2018), this article focuses on the case with binary response data, but the proof can be extended to the case where items are polytomous. The remainder of this section introduces the RHMM framework and presents a new identifiability theorem for RHMMs specific to the counterbalancing design.

2.3.1. General model

Let there be G total groups where each group corresponds to a unique ordering of the booklets. Observe that, for a given time point, two groups may have different booklets. That is, examinees from separate

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

groups may answer multiple different items at varying time points, resulting in different responses. In the HMM framework, this implies that the emission matrix $\mathbf{B}$ may change over time and we can define the multi-item emission matrix.

Definition 1 (Multi-item emission matrix). Suppose there are J total items with M_j representing the number of response options for item j. Let $\mathbb{J}_t \subset [J]$ to be the set of items administered at time t. The emission matrix of an item is defined as

$$\underset{M_j \times r}{\mathbf{B}_j} = \begin{bmatrix} b_{j,\,01} & \cdots & b_{j,\,0r} \\ \vdots & \cdots & \vdots \\ b_{j,\,M_{j-1},\,1} & \cdots & b_{j,\,M_{j-1},\,r} \end{bmatrix},$$

and the emission matrix at time t is

$$\underset{(\prod_{j\in\mathbb{J}_t} M_j)\times r}{\mathbf{B}_t} = \circledast_{j\in\mathbb{J}_t} \mathbf{B}_j,$$

where $\circledast_{j\in\mathbb{J}_t}$ denotes the Khatri–Rao product (i.e., a column-wise Kronecker product) of matrices $\mathbf{B}_j$ for all $j \in \mathbb{J}_t$ (see Definition A.3 in Appendix A).

Let K be the total number of binary latent attributes and J be the total number of items. The conditional distribution of the response $\boldsymbol{X}$ in the emission matrix is written in terms of $\boldsymbol{\alpha} \in \{0, 1\}^K$, as opposed to $\boldsymbol{Z}$. That is, for the remainder of this section on counterbalancing, the number of latent states is $r = 2^K$ corresponding to all possible attribute profiles. Here, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T)^\top$ denotes the learning trajectory of individual i, where $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{Kt})$ represents the binary attribute profile at time t. The element $\alpha_{kt} = 1$ indicates that the respondent possesses attribute k for $k \in [K]$, and $\alpha_{kt} = 0$ otherwise. When the items are dichotomous, the emission probability conditional on the attribute profile $\boldsymbol{\alpha}_t$ is

$$P(\boldsymbol{X}_t = \boldsymbol{x}_t \mid \boldsymbol{\alpha}_t, \boldsymbol{\Omega}) = \prod_{j=1}^{\mathbb{J}_t} b_{j,\,x_t,\,\boldsymbol{\alpha}_t}^{y_{jt}} (1 - b_{j,\,x_t,\,\boldsymbol{\alpha}_t})^{1-y_{jt}}, \tag{1}$$

where $\boldsymbol{\Omega}$ represents the model parameters, $\mathbb{J}_t$ denotes the set of items administered at time t with elements denoted by $\mathbb{j}_1, \ldots, \mathbb{j}_{|\mathbb{J}_t|}$, and $b_{j,\,x_t,\,\boldsymbol{\alpha}_t}$ is an element of an item-specific matrix, which denotes the probability of a response x_t to item j at time t for a subject with attribute profile $\boldsymbol{\alpha}_t$.

Equation (1) is a general form of the emission probability that contains the response probabilities for all 2^K latent classes for each item. The current model is unrestricted in the sense that there are no constraints on the emission probabilities. This lack of constraint, however, makes the general model susceptible to the label switching problem, a well-known issue in latent class modeling (Jasra et al., 2005; Redner & Walker, 1984). Specifically, across different groups and time points, the latent classes associated with the items can vary in order or even differ entirely. However, this problem is mitigated in RHMMs, where some structure is introduced by an expert-specified $\boldsymbol{Q}$ matrix (see below) and imposed on the emission matrix.

The restrictions can be connected to alternative representations where CDMs are written as a mixture of generalized linear models (de la Torre, 2011; Henson et al., 2009; von Davier, 2008). The response probability for an item j is now modeled as

$$b_{j,\,1,\,\boldsymbol{\alpha}_t} = P(X_{jt} = 1 \mid \boldsymbol{\alpha}_t, \boldsymbol{\beta}_j) = \Psi(\boldsymbol{\alpha}_t^{*\top} \boldsymbol{\beta}_j), \tag{2}$$

where $\Psi(\cdot)$ is an arbitrary cumulative distribution function,

$$\boldsymbol{\alpha}_t^{*\top} = \left(1, \alpha_{1t}, \ldots, \alpha_{Kt}, \alpha_{1t}\alpha_{2t}, \ldots, \alpha_{(K-1)t}\alpha_{Kt}, \ldots, \prod_{k=1}^{K} \alpha_{kt}\right) \tag{3}$$

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

is an alternative form of the binary vector $\boldsymbol{\alpha}_t$, and

$$\boldsymbol{\alpha}_t^{*\top}\boldsymbol{\beta}_j = \beta_{j,0} + \sum_{k=1}^{K}\beta_{j,k}\alpha_{kt} + \sum_{k>k'}\beta_{j,kk'}\alpha_{kt}\alpha_{k't} + \cdots + \beta_{j,12\ldots K}\prod_{k=1}^{K}\alpha_{kt}. \tag{4}$$

The regression coefficients create a sparse vector $\boldsymbol{\beta}_j$, where the nonzero elements represent the effects of latent skills or combination of skills on the response of item j. It can be shown that CDMs can be reparameterized to Equation (2) with different sparsity patterns of the vector $\boldsymbol{\beta}_j$. Following notations from Chen et al. (2020), let $\boldsymbol{\delta}_j = (\delta_{j0}, \ldots, \delta_{j,2^K-1})^\top \in \{0,1\}^{2^K}$ be the structure vector which indicates the relevant main effects and/or interactions of the K attributes for item j. The elements $\delta_{jc} = 0$ correspond to $\beta_{jc} = 0$ and are inactive, whereas $\delta_{jc} = 1$ implies that β_{jc} is active and non-zero. The matrix of structure vectors for a given time is denoted by $\Delta_t = (\boldsymbol{\delta}_{\mathbb{j}_1} \ \ldots \ \boldsymbol{\delta}_{\mathbb{j}_{|\mathbb{J}_t|}})^\top$. The structure matrices impose restrictions on the emission probabilities by providing information about the relationship between items and attributes. Using this general model to describe CDMs, we have the following new identifiability theorem for RHMMs in the context of a counterbalancing design.

Theorem 3 (Identifiability for RHMMs in a counterbalancing design). *Suppose there are J items, K attributes, T time points, G groups, and L booklets. Let $\mathbb{J}_t \subset [J]$ be the set of items administered at time t and $\boldsymbol{\pi}_1 = (\pi_{1,1}, \ldots, \pi_{1,2^K})$ is the initial distribution. For an HMM where time-varying emission matrices $\mathbf{B}_t$ contain probabilities formed by the general model in Equation (2) and $T \geq 3$, the parameters are identifiable if:*

(a) $rank(\mathbf{A}) = 2^K$;

(b) *for all $t \in [T]$, $\boldsymbol{\Delta}_t$ has the form $\boldsymbol{\Delta}_t = \begin{pmatrix} \boldsymbol{D} \\ \boldsymbol{\Delta}' \end{pmatrix}$ after row swapping, where $\boldsymbol{\Delta}'_t$ is any $|\mathbb{J}_t| \times 2^K$ binary matrix and $\boldsymbol{D}$ is a $K \times 2^K$ binary matrix with the structure*

$$\boldsymbol{D} = \begin{pmatrix} 1 & 1 & 0 & \ldots & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 1 & 0 & 0 & \ldots & 1 & \ldots & 0 \end{pmatrix},$$

which includes a diagonal pattern of 1's for the main-effect coefficients.

(c) *$G \geq L$ and there is a $t \in (1, T)$ such that all booklets show up once across G groups;*

(d) *$\pi_{1,l} > 0$ for all $l \in [2^K]$.*

Proof. The proof is found in Appendix B. □

To clarify Part (c), observe, for instance, time point 3 in Table 1. The design satisfies Part (c) because all five booklets appear once across the different groups. On the other hand, consider the design in Table 2. This design does have counterbalancing but it does not satisfy condition (c) in Theorem 3 because none of the time points in Table 2 contains all four booklets across the groups.

Part (d) of Theorem 3 requires the initial distribution to be strictly positive. That is, the samples are representative and include observations from all attribute profiles. Parts (a) and (b) assume that the emission and transition matrices are full rank. These are analogous to the requirement of a full-rank design matrix in linear regression. In practice, **A** and **B** may fail to be full rank when certain latent states are indistinguishable in their transition probabilities and response processes. An example for why this may happen is that the quality of items is bad such that they are unable to differentiate between the attribute profiles. Moreover, it should be noted that Part (b) is a sufficient condition for identifying model parameters and weaker conditions can be imposed for more general models (see, e.g., Culpepper, 2023). Hence, different CDMs and RLCMs may require different assumptions for identifiability of the

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Table 2. Unidentifiable design for $T = 4$

	Time points			
Groups	1	2	3	4
Group 1	Booklet 2	Booklet 1	Booklet 3	Booklet 4
Group 2	Booklet 4	Booklet 2	Booklet 3	Booklet 1
Group 3	Booklet 4	Booklet 3	Booklet 2	Booklet 1
Group 4	Booklet 3	Booklet 1	Booklet 4	Booklet 2

parameters. We will consider the case with the DINA as an example given that the DINA was used in Wang et al. (2018).

2.3.2. DINA model

We now introduce another binary matrix, which is more commonly used with the DINA, that also defines the underlying structure between the latent skills and items.

Definition 2 ($\boldsymbol{Q}$ matrix). Suppose there are J total items and K attributes. A $\boldsymbol{Q}$ matrix is a $J \times K$ binary matrix that can be written as $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_J)^\top \in \{0,1\}^{J \times K}$. The $\boldsymbol{q}_j^\top = (q_{j1}, \ldots, q_{jK})$ represents the j-th row of the Q-matrix. Element $q_{jk} = 1$ indicates that answering item j requires the mastery of skill k while $q_{jk} = 0$ means that skill k is not required for item j.

It is possible to derive the $\boldsymbol{Q}$ matrix when the structure matrix is known. Different from $\boldsymbol{\Delta}_t$, the $\boldsymbol{Q}$ matrix only contains partial information about the item–attributes relationship (Fang et al., 2019). The $\boldsymbol{Q}$ matrix does not provide information on how the attributes may interact with each other without specifying a particular CDM. Despite this limitation, the $\boldsymbol{Q}$ matrix is widely used in CDM research, and its identifiability has been extensively explored (Chen et al., 2015; Gu & Xu, 2019; Xu, 2017; Xu & Shang, 2018). We consider the DINA model along with the assumption of a known $\boldsymbol{Q}$ matrix.

For a DINA, the probability of a correct response is

$$b_{j,\,1,\,\boldsymbol{\alpha}_t} = P(X_{jt} = 1 \mid \boldsymbol{\alpha}_t, \boldsymbol{q}_t, s_j, g_j) = (1 - s_j)^{\eta_{jt}} g_j^{1-\eta_{jt}}, \tag{5}$$

where $\eta_{jt} = 1$ if an individual mastered the required attributes for item j at time t and $\eta_{jt} = 0$ if at least one required attribute is not mastered. Additionally, we have $s_j = P(X_{jt} = 0 \mid \eta_{jt} = 1)$ and $g_j = P(X_{jt} = 1 \mid \eta_{jt} = 0)$ correspond to the slipping and guessing parameter of an item j. Now, Theorem 3 can be rewritten for the current setting with the DINA as

Corollary 1 (Identifiability for RHMMs-DINA in a counterbalancing design). *Suppose there are J items, K attributes, T time points, G groups, and L booklets. For an HMM where time-varying emission matrices $\mathbf{B}_t$ contain probabilities formed by the DINA model shown in Equation (5) and $T \geq 3$, the parameters are identifiable if:*

(a) $rank(\mathbf{A}) = 2^K$;
(b) $\mathbf{Q}_t$ *takes the form (after a row permutation):*

$$\mathbf{Q}_t = \begin{pmatrix} \boldsymbol{I}_K \\ \mathbf{Q}_t^* \end{pmatrix}_{J \times K}$$

with $\boldsymbol{I}_K$ as a $K \times K$ identity matrix and $\mathbf{Q}_t^*$ can be any form except zero rows;
(c) $G \geq L$ *and there exists $t \in (1, T)$ such that all booklets show up once across G groups;*
(d) $\pi_{1,\,l} > 0$ *for all $l \in [2^K]$.*

Proof. The proof is found in Appendix C. □

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Remark 1. Part (b) in Corollary 1 is specific to the DINA case and corresponds to Part (b) in Thereom 3. It is worth noting that this condition is related to the reliance of the proof on Kruskal's theorem, as it imposes some requirements on $\boldsymbol{Q}_t$ when the DINA model is used (see Remark C.1 in Appendix C). Moreover, Part (b) in Corollary 1 also indicates that the $\boldsymbol{Q}_t$ matrix at each time point is complete (Chiu et al., 2009), which is a common condition for identifying the DINA model when there is only one time point (Gu & Xu, 2019). The completeness condition guarantees that the $\boldsymbol{Q}_t$ matrix can identify all potential attribute patterns.

The identifiability theorems introduced thus far are limited to RHMMs. There are certain advantages to using RHMMs over conventional HMMs. Having either the $\boldsymbol{Q}$ or the structure matrix, some information can be obtained regarding the underlying structure of the latent states to help reduce the number of estimated parameters. It is worth noting that the identifiability conditions in Thereom 3 and Corollary 1 do not require the $\boldsymbol{Q}$ matrix or the structure matrix to be known but rather to take certain form. Whether or not this assumption is feasible is beyond the scope of this article; however, there has been substantial research on estimating the $\boldsymbol{Q}$ matrix more accurately (Balamuta & Culpepper, 2022; Chen et al., 2018, 2020; Köhn et al., 2025).

Another benefit of using RHMMs, as mentioned briefly before, is that using RHMMs with known $\boldsymbol{Q}$ matrices avoids the issue of label switching by mapping the responses $\boldsymbol{X}$ onto a more parsimonious collection of binary attributes. In conventional HMMs, the time-varying emission matrices are susceptible to the label switching problem. Specifically, for different groups, $\mathbf{B}_t$ can represent different orderings of the latent classes, or even entirely different underlying latent states across groups. Methods that enforce unique labeling across groups are thus necessary to ensure meaningful interpretations and inference.

2.3.3. Empirical investigation of counterbalanced design identifiability

Theorem 3 establishes sufficient conditions for identifying the parameters of RHMMs within a counterbalancing research design. Part (c) is particularly relevant to the research setting for which we want to provide additional insights. Wang et al. (2018) offer intuition on how the counterbalancing helps with identifying the DINA model parameters by exposing the items to individuals with varying mastery attribute profiles. This corresponds to the assumption of nonzero population proportions (Part (d) of Theorem 3). Moreover, counterbalancing addresses potential confounding factors such as test fatigue and systematic changes in item difficulty over time that are unrelated to changes in individuals' mastery. For instance, one might observe the items administered at later time points appear more difficult (e.g., lower guessing parameters in the DINA model). Without counterbalancing, it would be unclear whether this phenomenon reflects test fatigue, genuine increases in item difficulty, proficiency loss, or some combinations of these factors. Consequently, the lack of counterbalancing is likely to introduce potential issues with parameter recovery of RHMMs.

To demonstrate the potential consequences of failing to meet the identifiability conditions, we conducted a simple simulation. The simulation design followed Wang et al. (2018) and included 350 individuals, 50 items, and 5 time points. We employed the same Q-matrix from Wang et al. (2018) and used the DINA first-order HMM (DINA-FOHM; Chen et al., 2018) to generate the data. The DINA-FOHM is conceptually similar to the model described in the previous section which integrates DINA with RHMM. Data were generated using the HMCDM package (Zhang et al., 2024). Two sets of item parameters were generated, which we call the *normal* and *easier* conditions. In the normal condition, the slipping and guessing parameters for all items were drawn from a $U(0.1, 0.2)$ distribution. In the easier condition, the guessing parameters for the items increased over time by generating from uniform distributions with progressively higher bounds. This condition accounts for a potential confound where improvements in performance are not solely a result of changes in students' mastery levels. Additionally, we manipulated the research design such that there exists counterbalancing in one condition, similar to Wang et al. (2018) and Table 1. The other situation does not have counterbalancing, so that all individuals receive same booklet of items at each time point and the item parameters change over time. Only one replication was conducted for each condition.

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

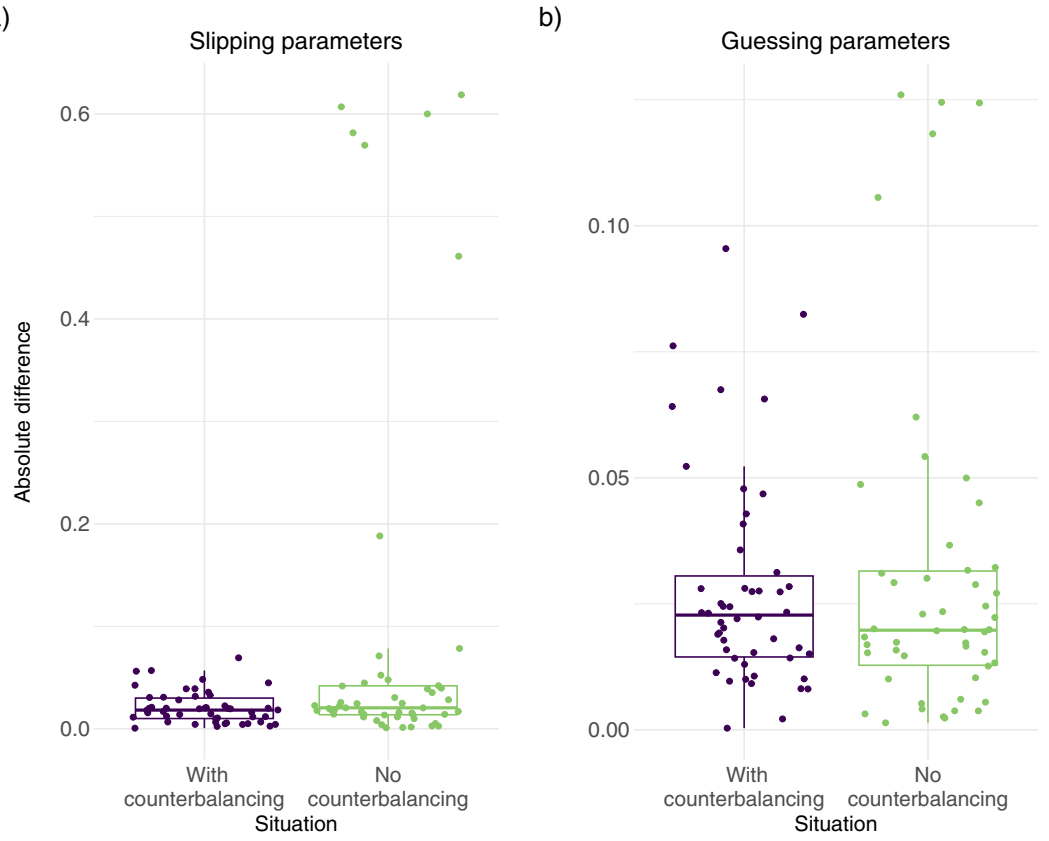


Figure 1. Absolute difference between the estimated and true item parameters in the normal case (similar items over time).

We estimated the item parameters using the generated data and computed the absolute difference between the estimated and true item parameters. Figure 1 presents the results for the normal condition. Observe that with counterbalancing, the item parameters were accurately recovered. On the other hand, when no counterbalancing is present, parameter recovery was poorer, particularly for some items where the absolute differences in the slipping parameters are around 0.6. This result is even more pronounced for the easier condition in Figure 2. Without counterbalancing, several items exhibited large absolute differences for both slipping and guessing parameters. With counterbalancing, the item parameters remained well estimated/recovered. From this small simulation, we argue that it is crucial to consider model's assumptions to ensure identifiability and to enable meaningful interpretation of the model's results.

2.4. Multiple-group longitudinal design

Another commonly employed experimental design is the multiple-group longitudinal design. This design is widely used to evaluate the effectiveness of interventions across various domains, including psychology (van Hoorn et al., 2016), education (Madison & Bradshaw, 2018b; Tang & Zhan, 2021), and clinical trials (Khori et al., 2015). The design serves as a tool for investigating whether the effects observed can be attributed to the intervention itself rather than to extraneous factors. Figure 3 shows the research flowchart of two consecutive time points within the multiple-group longitudinal design when there are three groups. Individuals are randomly assigned to treatment and control groups and exposed to different interventions depending on the group membership. All groups of individuals take the

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

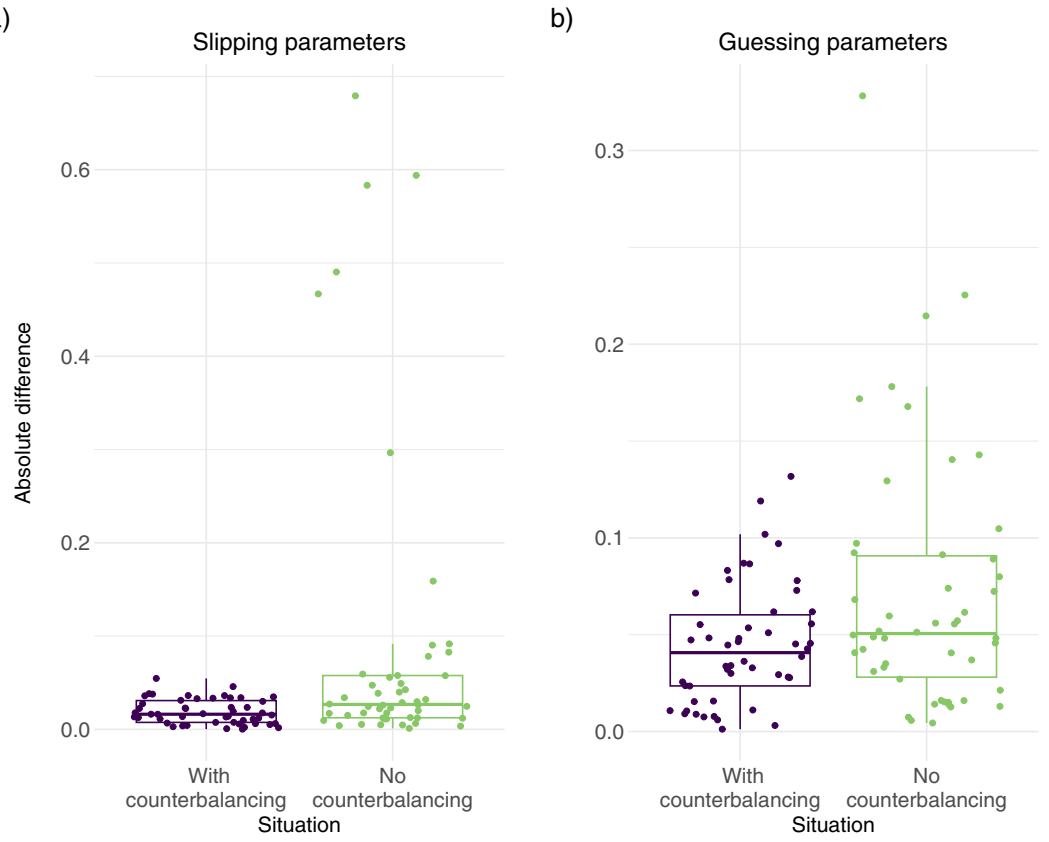


Figure 2. Absolute difference between the estimated and true item parameters where items get easier over time.

same assessment at time $t-1$. Conceptually, this design is somewhat similar to the multi-armed bandit problem in reinforcement learning where there are multiple options with different reward distributions. In contrast to adaptive bandit algorithms, the experimental design fixes group membership at the outset as individuals are randomly assigned to a single intervention arm and remain in that group for the duration of the study. The control group functions as a baseline and typically receives a placebo or no intervention which should yield minimal to no improvement compared to the treatment groups. Changes in the performance are tracked over time both within each group and between groups. The intervention(s) is considered to be effective when the changes observed in the treatment group(s) are determined to be practically and/or statistically better than that of the control group.

Consider, for example, the experiment in Tang and Zhan (2021) that assessed the effectiveness of an intervention called CDF in promoting learning. The authors compared this feedback mode to two other interventions: correct–incorrect response feedback (CIRF) and no feedback. To implement this, they used a quasi-experimental design in which different classes within a school were randomly assigned to one of the feedback types. Students took three parallel tests at different time points, with the interventions applied between testing occasions. The data were analyzed using RHMMs but the identifiability was not investigated. Using this example as a guide, we adapt a more general HMM for this design and derive an identifiability theorem for HMM in this context.

In the experiment by Tang and Zhan (2021), the control group consisted of individuals who did not receive any feedback after a test. Individuals in the control group are assumed to exhibit no improvement at all relative to those in the treatment groups. In this context, the assumption is equivalent to fixing the transition matrix of the control group to be the identity matrix. In contrast, one would expect the

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

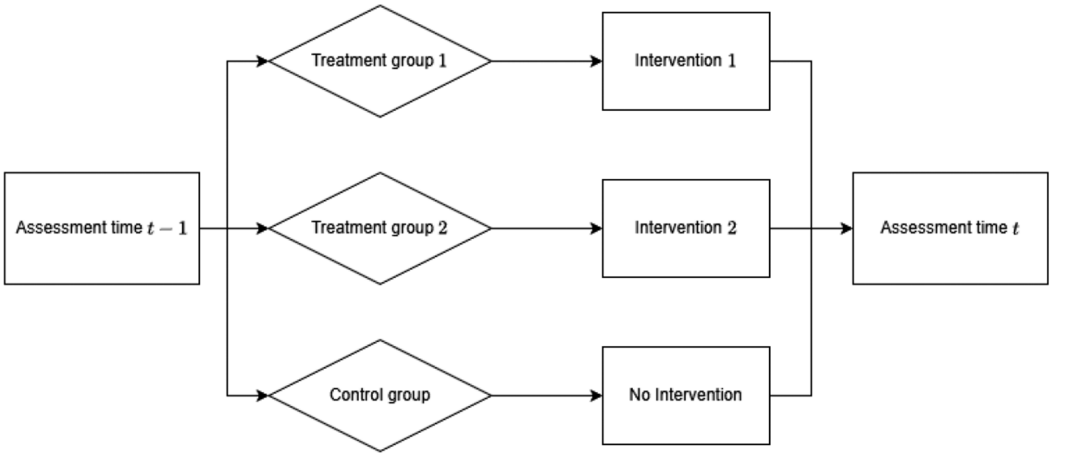


Figure 3. Flowchart for two consecutive points of the multiple-group longitudinal design.

treatment groups to exhibit changes in the latent attributes. Their transition matrices would differ not only from the identity matrix, but potentially also from each other. Unlike the counterbalancing design, individuals from different groups answer the same items at each time point. However, the items can differ between time points. Therefore, we still consider an HMM where the emission matrix $\mathbf{B}_t$ changes over time and the transition matrix $\mathbf{A}$ differs between groups.

Following the same setting in Section 2.1, suppose there are N subjects, J items, T time points, and r latent states. Let $\mathbf{A}^{(g)}$ represent the transition matrix for group g with $\mathbf{A}^{(c)}$ denote the transition matrix for the control group. We have the following theorem for HMM identifiability with respect to this multiple-group longitudinal design.

Theorem 4 (Identifiability for HMMs in a multiple-group longitudinal design). *Suppose there are J items, T time points, and r latent states. For an HMM with time-varying emission matrix $\mathbf{B}_t$, group-specific initial distribution $\boldsymbol{\pi}_1^{(g)} = (\pi_{1,1}^{(g)}, \ldots, \pi_{1,r}^{(g)})$, group-specific transition matrix $\mathbf{A}^{(g)}$, and $T \geq 3$, the parameters are identifiable if:*

(a) *$rank(\mathbf{A}^{(g)}) = r$ for all group g;*
(b) *$rank(\mathbf{B}_t) = r$ for all $t \in [T]$;*
(c) *$\mathbf{A}^{(c)} = \boldsymbol{I}_r$;*
(d) *$\pi_{1,l}^{(g)} > 0$ for all $l \in [r]$ across all group g.*

Proof. The proof is found in Appendix D. □

2.4.1. Empirical investigation of multigroup longitudinal design identifiability

Similar to the counterbalancing design, we now want to present additional clarification for the identifiability conditions in Theorem 4. Specifically, we focus on Part (c), its interpretation, and how it facilitates the identification of HMM parameters. Part (c) assumes that individuals in the control group do not exhibit any changes in their latent classes. In essence, the control group serves as an anchor that, similar to the use of anchor items to fix the scale of item parameters in item response theory, ensures identification of both the parameters and latent classes. The existence of a control group helps to address the potential issue of label switching that arises when employing a Bayesian framework to estimate HMM parameters (Jasra et al., 2005). Additionally, the inclusion of a control group mitigates potential confounding factors introduced when different items are administered across time points. Without a control group with constant latent states, it might be difficult to evaluate whether changes in

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Table 3. Identifiability results for different hidden Markov models

Existing literature	Constant emission	Latent structure	Group-specific transition
Allman et al. (2009)	Yes	No	No
Bonhomme et al. (2016)	Yes	No	No
Liu et al. (2023)	Yes	Yes & No	No
Liu and Culpepper (2025)	No	Yes & No	No
Current work			
Theorem 3 (CB)	No	Yes	No
Corollary 1 (CB–DINA)	No	Yes	No
Theorem 4 (Multiple-group)	No	No	Yes

Note: CB = Counterbalancing. Latent structure assumption indicates whether the emission probabilities are restricted, distinguishing RHMMs from HMMs.

the estimated item parameters between time points are due to different mastery levels and/or genuine differences in item parameters.

We again performed a simple simulation to illustrate what would happen when the identifiability conditions are not met. The simulation design consisted of 350 individuals, 30 items, and 3 time points. Consistent with the previous simulation settings, the data were generated under the DINA-FOHM model. More specifically, a single dataset was simulated with a control group whose latent classes remained constant over time. Using this dataset, we evaluated the role of the control group by comparing two scenarios: one in which the software incorporated the control group information, and another where the software did not. In the first condition, all individuals answered the same set of 10 items at each time point. As existing algorithms do not allow the specification of individuals' group membership, part (c) in Theorem 4 is violated, because all individuals are implicitly assumed to have a nonzero probability of transitioning to different latent classes over time. To reflect the notion of a non-changing control group within the constraints of current algorithms, one possible approach is to require those in the control group to answer all items at the first time points, rather than administering sets of 10 items across three time points. Hence, in the second condition, the existence of a control group is specified by having a number of individuals answer all 30 items at the first time point. Following the setup in Section 2.3.3, only one replication was conducted for each condition.

Data were fitted using the HMCDM and TDCM (Madison et al., 2025) packages, and the absolute difference between the estimated and true item parameters was recorded. The results are illustrated in Figure 4. When there is a control group, the item parameters were accurately recovered. On the other hand, without a control group, parameter recovery was poorer for both the slipping and guessing parameters. These findings further support the necessity to ascertain that the conditions for identification of models' parameters are satisfied.

3. Practical suggestions for practitioners

This section provides a summary of the aforementioned results, along with existing research, for practitioners who are planning to design assessments within the HMM or RHMM frameworks. Table 3 outlines some of the considerations for selecting the appropriate results to guide study design.

Table 3 highlights three main design considerations. The first concerns whether the emission matrix is constant over time. To determine whether $\mathbf{B}_t$ varies, researchers should evaluate whether the instrument administered exhibits invariant response probabilities. In educational settings, this can be

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

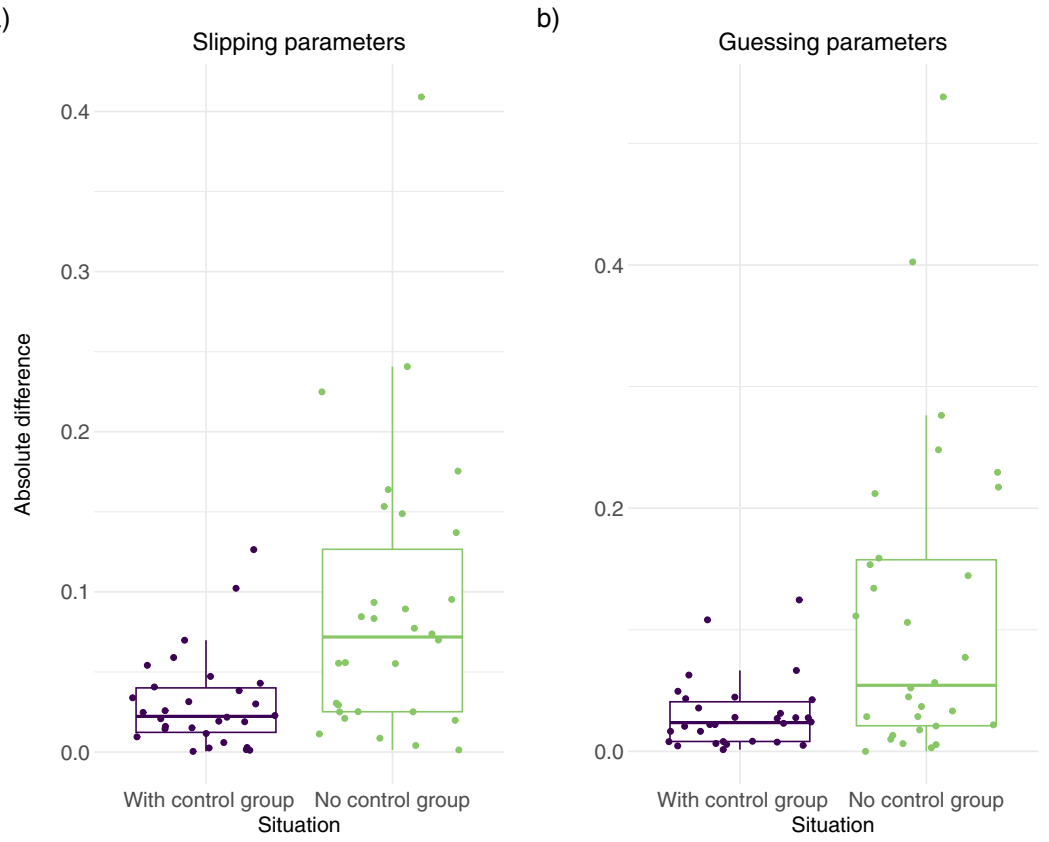


Figure 4. Absolute difference between the estimated and true item parameters for the multiple-group design with a control group.

represented by a set of common or parallel items that require mastery of the same attributes. When there are different items administered at different time points, we should expect the emission matrices to change. For a constant emission matrix, we refer readers to Allman et al. (2009), Bonhomme et al. (2016), and Liu et al. (2023). In contrast, the conditions in the current article, as well as in Liu and Culpepper (2025), allow for the emission matrix to vary over time.

The second consideration relates to whether the emission probabilities are restricted by a prespecified structure linking latent skills and items. If there exists a structure, the emission probabilities can be modeled using RHMMs with appropriate RLCMs. The relevant identifiability results to guide study design in this scenario are in Theorem 3 and Corollary 1, as well as in prior work (Liu & Culpepper, 2025; Liu et al., 2023). Note that Theorem 3, Corollary 1, and results in Liu and Culpepper (2025) allow $\mathbf{B}_t$ to vary over time, whereas Liu et al. (2023) do not. The results in the current article differ from Liu and Culpepper (2025) in terms of the underlying experimental design. Liu and Culpepper (2025) require at least one pair of adjacent time points with parallel items while no such requirement is imposed here. For general HMMs, the identifiability results are stated in Theorem 4 together with the existing research cited in Table 3. Similar to the present work, Liu et al. (2023) and Liu and Culpepper (2025) establish conditions for both RHMMs and HMMs.

The last point focuses on whether the transition matrix should differ across groups. This is related to situations where there may be multiple groups with heterogeneous transition parameters. Group-specific transition matrices can provide additional information to evaluate whether between-group interventions influence how the latent states change over time. To our knowledge, Theorem 4 is currently the only identifiability result that accommodates this scenario. Specifically, we require the presence of

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

a control group that is assumed to be unchanged over time. Moreover, the initial distributions for each group must be strictly positive, implying that practitioners must ensure their samples are representative and include respondents from all latent states.

4. HMM parameter estimation for multigroup longitudinal design

From Theorem 4, we introduce a Bayesian framework and Gibbs sampling algorithm to estimate the model parameters for a conventional HMM within the setting of a multiple-group longitudinal design. Suppose that there are N subjects, J items, T time points, and r latent states. At time t, each group receives the same items and we have the response probabilities of an individual i:

$$P(\boldsymbol{X}_{i,t} \mid Z_{i,t}, \mathbf{B}_1, \ldots, \mathbf{B}_{\mathbb{J}_t}) = \prod_{j \in \mathbb{J}_t} P(X_{i,j,t} \mid Z_{i,t}, \mathbf{B}_j) = \prod_{j \in \mathbb{J}_t} b_{j,x_{i,t},z_t},$$

where $\mathbb{J}_t$ is the set of items administered at time t and $\mathbf{B}_j$ is the emission matrix for item j. The likelihood of observing a sample of N responses to J items with T time points is defined as

$$P(\boldsymbol{X} \mid \boldsymbol{Z}, \mathbf{B}) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{z=1}^{r} \left[\prod_{j \in [J]} \left(\prod_{x \in [M_j]} b_{j,x,z}^{\mathbb{1}(x_{i,t}=x)} \right)^{\mathbb{1}(j_{i,t}=j)} \right]^{\mathbb{1}(z_{i,t}=z)}, \tag{6}$$

where M_j is the number of response options for item j and $\mathbb{1}(\cdot)$ is the indicator function. By aggregating over individuals and time points, we get

$$P(\boldsymbol{X} \mid \boldsymbol{Z}, \mathbf{B}) = \prod_{j \in [J]} \prod_{x \in [M_j]} \prod_{z \in [r]} b_{j,x,z}^{n_{j,x,z}}, \tag{7}$$

where $n_{j,x,z} = \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{1}(x_{i,t} = x)\mathbb{1}(j_{i,t} = j)\mathbb{1}(z_{i,t} = z)$ denotes the number of individuals within class z whose responses to item j are x.

The model for the latent state of an individual i is an r-dimensional categorical distribution that depends on the time point t and the person's group membership. Recall that $\mathbf{A}^{(g)}$ and $\boldsymbol{\pi}_1^{(g)}$ are the transition matrix and initial distribution of a group g, we can write the conditional density for $Z_{i,t}$

$$P(Z_{i,t} = z_{i,t} \mid Z_{i,t-1} = z_{i,t-1}, \mathbf{A}^{(g)}, \boldsymbol{\pi}_1^{(g)}) = \begin{cases} \pi_{1,z_{i,1}}^{(g)}, & t = 1 \\ a_{z_{i,t-1},z_{i,t}}^{(g)}, & t > 1, \end{cases} \tag{8}$$

where $a_{z,z'}^{(g)}$ is an element in $\mathbf{A}^{(g)}$. Then, the joint conditional distribution of latent states for N responses is a product over time, observations, and groups. That is, let there be G total groups and suppose $\mathbf{A}$ and $\boldsymbol{\pi}_1$ now represent the collection of the transition matrix and initial distributions for all G group. The joint conditional distribution of $\boldsymbol{Z}$ for all individuals is

$$P(\boldsymbol{Z} \mid \mathbf{A}, \boldsymbol{\pi}_1) = \prod_{i=1}^{N} \left[\left(\pi_{1,z_{i,1}}^{(g_i)} \right) \times \prod_{t=2}^{T} \left(a_{z_{i,t-1},z_{i,t}}^{(g_i)} \right) \right], \tag{9}$$

where g_i denotes the group membership of individual i and $g_i \in [G]$. It is important to note that the transition matrix of the control group is assumed to be the identity matrix, which corresponds to the fact that the control group does not exhibit changes in the latent states.

For any group g, let $\boldsymbol{X}^{(g)}$ and $\boldsymbol{Z}^{(g)}$ be the observations and latent states of all individuals within the group. The posterior distribution of the parameters for the HMM is

$$P(\boldsymbol{Z}^{(g)}, \mathbf{B}, \mathbf{A}^{(g)}, \boldsymbol{\pi}_1^{(g)} \mid \boldsymbol{X}^{(g)}) \propto P(\boldsymbol{X}^{(g)} \mid \mathbf{B}, \boldsymbol{Z}^{(g)}) P(\boldsymbol{Z}^{(g)} \mid \boldsymbol{\pi}_1^{(g)}, \mathbf{A}^{(g)}) P(\mathbf{B}) P(\boldsymbol{\pi}_1^{(g)}) P(\mathbf{A}^{(g)}).$$

To obtain the priors for the parameters, we first write $\mathbf{B}_j = (\boldsymbol{b}_{j,1}, \ldots, \boldsymbol{b}_{j,r})$ such that column $\boldsymbol{b}_{j,z} \in \mathcal{S}_{M_j}$, where $\mathcal{S}_k = \{\boldsymbol{x} \in \mathbb{R}^k : \boldsymbol{x}^\intercal 1_k = 1, x_i \geq 0 \, \forall \, i \in [k]\}$ is the n-dimensional simplex. Similarly, let $\mathbf{A}^{(g)} =$

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

$(\boldsymbol{a}_1^{(g)}, \ldots, \boldsymbol{a}_r^{(g)})^\top$ such that row $\boldsymbol{a}_z^{(g)} \in \mathcal{S}_r$. Using independent Dirichlet distributions, we have

$$\{\boldsymbol{b}_{j,z}\}_{z=1}^{r} \sim \text{Dirichlet}(\boldsymbol{d}_{\mathbf{B}_j}) \tag{10}$$

$$\{\boldsymbol{a}_z^{(g)}\}_{z=1}^{r} \sim \text{Dirichlet}(\boldsymbol{d}_{\mathbf{A}^{(g)}}) \tag{11}$$

$$\boldsymbol{\pi}_1^{(g)} \sim \text{Dirichlet}(\boldsymbol{d}_{\boldsymbol{\pi}_1^{(g)}}), \tag{12}$$

where $\boldsymbol{d}_{\mathbf{B}_j}$, $\boldsymbol{d}_{\mathbf{A}^{(g)}}$, $\boldsymbol{d}_{\boldsymbol{\pi}_1^{(g)}}$ can be set depending on the prior used. For instance, a Jeffrey's prior can be applied to the simplex by setting $\boldsymbol{d}_{\mathbf{B}_j} = 1_{M_j}/2$, $\boldsymbol{d}_{\mathbf{A}^{(g)}} = 1_r/2$, $\boldsymbol{d}_{\boldsymbol{\pi}_1^{(g)}} = 1_r/2$. The above priors assume independent Dirichlet distributions for the columns of the emission matrix and for the rows of the group-specific transition matrix. We can then obtain the following priors for $\mathbf{B}$, $\mathbf{A}^{(g)}$, and $\boldsymbol{\pi}_1^{(g)}$:

$$P(\mathbf{B}) \propto \prod_{z\in[r]} \prod_{j\in[J]} \prod_{x\in[M_j]} b_{j,x,z}^{d_{\mathbf{B}_j,z}-1} \mathbb{1}(\boldsymbol{b}_{j,z} \in \mathcal{S}_{M_j}) \tag{13}$$

$$P(\mathbf{A}^{(g)}) \propto \prod_{z\in[r]} \prod_{z'\in[r]} \left(a_{z,z'}^{(g)}\right)^{d_{\mathbf{A}^{(g)},z'}-1} \mathbb{1}(\boldsymbol{a}_z^{(g)} \in \mathcal{S}_r) \tag{14}$$

$$P(\boldsymbol{\pi}_1^{(g)}) \propto \prod_{z\in[r]} \left(\pi_{1,z}^{(g)}\right)^{d_{\pi_{1,z}^{(g)}}-1} \mathbb{1}(\boldsymbol{\pi}_1^{(g)} \in \mathcal{S}_r). \tag{15}$$

The Gibbs sampling algorithm is employed to sample from the posterior distribution. Full conditional distributions of the parameters are included in Appendix E.

5. Simulation study for evaluating parameter recovery performance

5.1. Settings

We next report results from a Monte Carlo experiment to evaluate the performance of the Gibbs sampling algorithm in recovering the parameters under the identification conditions of Theorem 4. We conducted the simulation study with three time points, and two groups: one treatment and one control. Other design factors that were manipulated include number of latent attributes (i.e., $K = 2, 4, 6,$ and 8), sample sizes per group (i.e., $N = 100, 500,$ and $1,000$), and number of items per time point (10, 20, and 100). For each setting, we repeated the simulation study 500 times.

By the identifiability condition, the transition matrix for the control group was specified as the identity matrix. For the treatment group, the transition matrix was simulated such that it is full rank and each row included entries with relatively higher probabilities compared to the others. This structure captures potential real-life situations where certain latent states might be more likely to transition into one another, indicating potential correlations among these states. Moreover, all elements of the treatment group's transition matrix were nonzero. That is, we allowed transitions between any pairs of states to demonstrate the flexibility of the model. In practical settings, researchers may evaluate whether this transition matrix is realistic and, if necessary, impose additional constraints.

The emission matrices for the items were simulated to be full rank across time points and allow for potential separation between the latent classes. Specifically, all items were dichotomous, and the probabilities of a correct response of an item are simulated from a Beta distribution. The beta parameter is fixed at 10 and the alpha parameter is a function of the latent classes. This corresponds to different latent classes having different means or locations. That is, the latent classes, on average, have different probabilities of answering a question correctly. The emission matrices across time points were generated from the same underlying distributions to reflect the use of parallel items similar to the conditions described in Tang and Zhan (2021). The initial distributions for the two groups were set to be uniform from all possible latent classes. That is, the initial distributions have strictly positive elements, which satisfies conditions (d) of Theorem 4.

Lastly, we utilized the identifiability assumption to find good initial values for the Gibbs sampling algorithm. Particularly, we performed a latent class analysis on the control group to estimate the

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

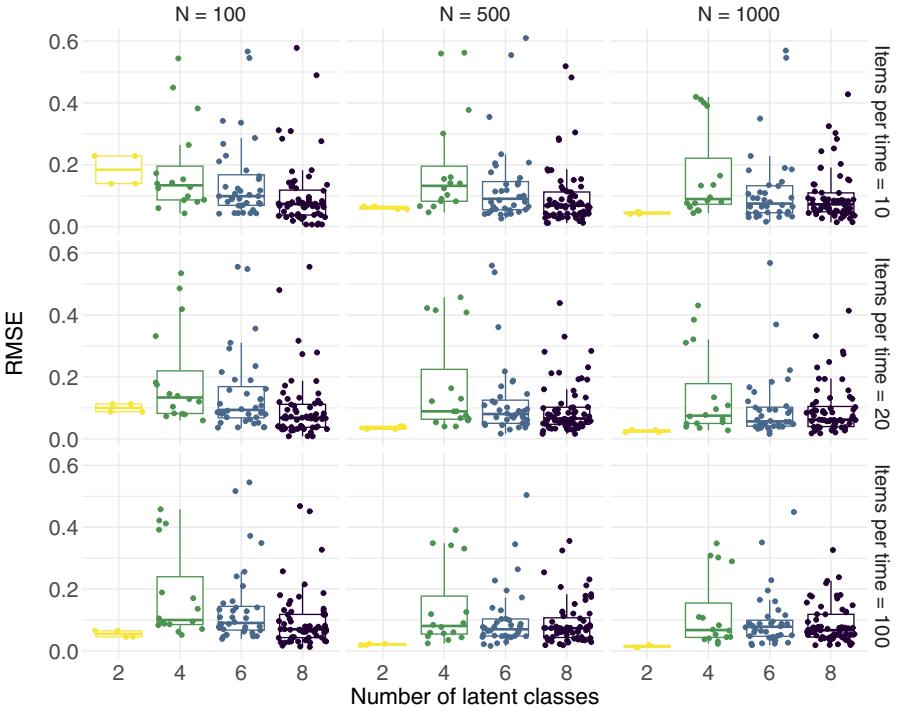


Figure 5. Boxplot of RMSEs of the transition matrices.

emission matrix. Based on the estimates, we identified the most likely latent classes for individuals in the treatment group at each time point. These classifications were then used to derive the transition matrix and initial distribution. Providing the algorithm with more informed starting values may help the Markov chain converge more quickly and, therefore, reduce the length of the burn-in period and overall computational burden. We used Markov chains of length 70, 000 with the first 15, 000 discarded as burn-in for all conditions.

5.2. Results

We computed the estimated transition and emission matrices by taking the mean of all samples after the burn-in period. Moreover, we estimated the latent classes for all individuals at each time point by selecting the most frequent class across all samples after the burn-in. Two main metrics were used to evaluate parameter recovery. We calculated the average root mean squared error (RMSE) to assess the accuracy of the estimated transition and emission matrices. For the classification, we provide the mean accuracy rate across all time points and replications by comparing the estimated with the true latent class.

Figures 5 and 6 show the RMSE values for the transition and emission matrices, respectively, across all conditions. Overall, the results suggest reasonably good recovery for the transition and emission matrices. The RMSE values increase as the number of individuals per group decreases and/or the number of latent classes increases. This pattern is expected as a greater number of latent classes corresponds to a larger number of parameters to estimate, thereby requiring more data for accurate recovery. The number of items per time point does not seem to affect the recovery rate of the transition and emission matrices. In Figure 5, the elements with high RMSE values are primarily associated with entries with large probabilities within a row. While the estimated transition matrices were generally able to identify these high-probability transitions, they tended to underestimate their true magnitudes.

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

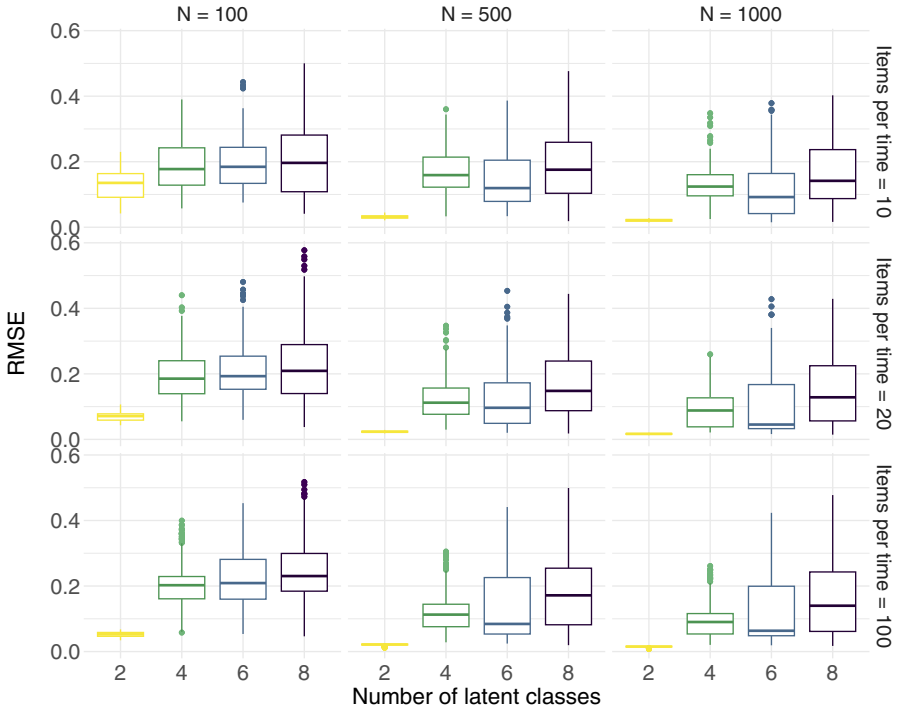


Figure 6. Boxplot of RMSEs of the emission matrices.
Note: RMSE = Root mean squared error.

Figure 7 illustrates the classification accuracy. Generally, the estimated latent class performs better than random guessing. For a fixed number of latent classes and number of individuals per group, the classification accuracy improves as the number of items per time point increases. This is expected given the availability of more information per individual. For $N = 100$, however, the improvement is less pronounced. One potential explanation for this result is that classification accuracy also depends on the performance of the estimated parameters. Note that increasing the number of individuals per group leads to better classification accuracy for different number of latent classes. Moreover, at $N = 1,000$, there is a clearer improvement in accuracy as the number of items per time point increases compared to when the sample size is 100. Again, this finding is likely associated with having more accurate estimation of the transition and emission matrices associated when the sample size increases.

6. Data application

The purpose of this section is to illustrate the application of the proposed algorithm to a real-world data analysis problem. The dataset contains responses to three sets of $J = 18$ items in rational number operations (Tang & Zhan, 2021). Six classes, with assumed similar academic level based on admission scores, were randomly assigned to three conditions. The conditions were based on the types of intervention students received in between the testing periods. The three conditions were: cognitive diagnostic feedback (CDF) group, CIRF group, and no feedback (control) group. The CDF group is referred to as the diagnosis group and the CIRF group is denoted as the traditional group. A total of 289 students participated but only responses from 276 students were included in the dataset. Specifically, dichotomous data of 90, 92, and 94 students were recorded for the diagnosis, traditional, and control groups, respectively. We compared models with the number of latent classes from $r = 2$ to $r = 8$ using

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

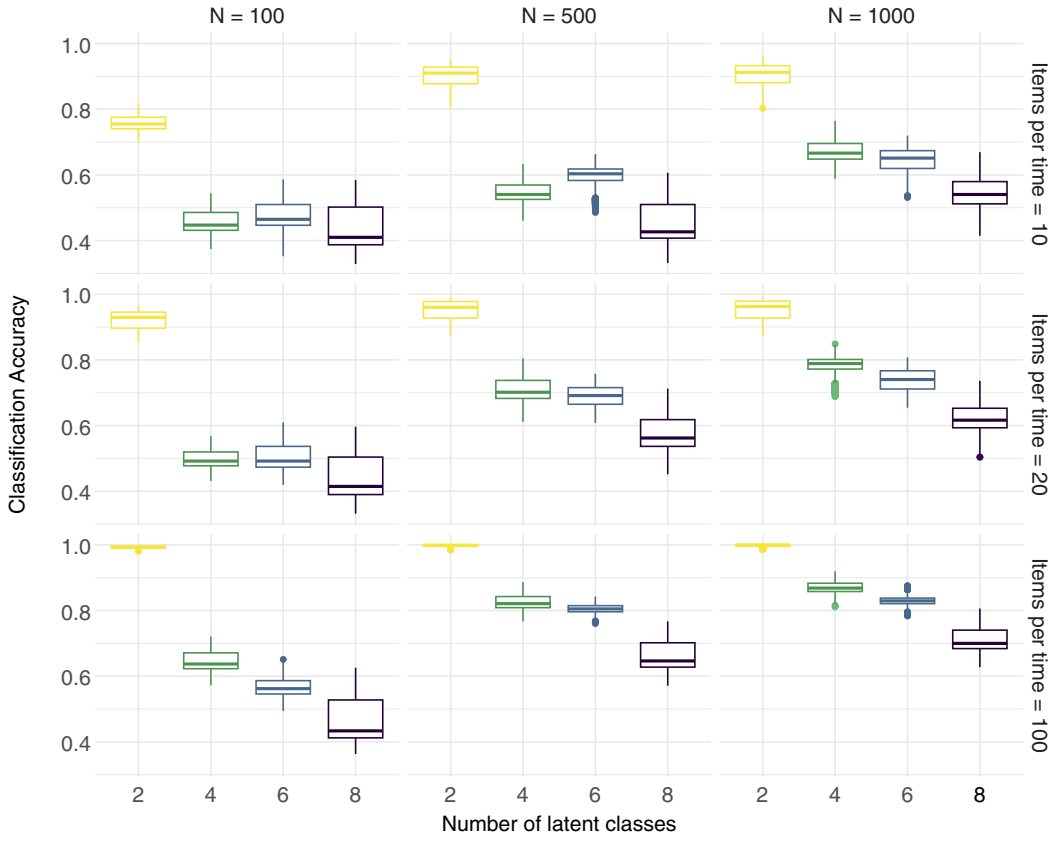


Figure 7. Boxplot of classification accuracy.

the leave-one-out information criterion (Vehtari et al., 2017). We set $r = 7$ based on the results. It should be noted that this number of latent classes would not be possible for an RHMM or any model which utilizes an RLCM. Therefore, this demonstrates the flexibility and potential of a conventional HMM.

An important contribution of the current work is that the proposed identifiability conditions permit direct comparisons across treatment groups under the HMM framework. Specifically, if a researcher is willing to make the assumptions in Theorem 4 (e.g., individuals in the control group remain in the same latent class over time, the items are good and can distinguish among the seven specified latent classes), then they can estimate group-specific transition probabilities. These probabilities can then be used to assess the effects of interventions on learning trajectories and to facilitate comparisons across groups. Tables 4 and 5 report the estimated transition matrix for the two treatment groups.[1] For instance, the first rows of Tables 4 and 5 show that students in state 1 tend to remain there or change to state 3. This might indicate that the CDF and CIRF interventions do not differ in their impact on students who are in state 1. Similarly, we can examine the effects of the interventions for students in different latent classes.

Additionally, the distribution of latent classes over time for a specific group can be calculated from the estimated initial distribution and transition matrices. The estimated distribution of latent classes for the traditional group is illustrated in Table 6 while that of the diagnosis group is depicted in Table 7. These distributions may provide general insight into the effects of the interventions. For this data application, the initial distributions of both groups are assumed to be the same. From the tables, it appears that state

[1]Throughout the section, values of probabilities will be rounded to the third decimal place

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Table 4. Estimated transition matrix for the traditional group

Latent class	1	2	3	4	5	6	7
1	0.358	0.024	0.504	0.070	0.016	0.014	0.014
2	0.032	0.321	0.063	0.502	0.027	0.027	0.027
3	0.085	0.029	0.371	0.458	0.020	0.019	0.019
4	0.021	0.062	0.021	0.861	0.011	0.011	0.012
5	0.037	0.035	0.040	0.036	0.777	0.036	0.039
6	0.019	0.019	0.019	0.019	0.093	0.811	0.020
7	0.013	0.013	0.013	0.013	0.049	0.247	0.654

Table 5. Estimated transition matrix for the diagnosis group

Latent class	1	2	3	4	5	6	7
1	0.245	0.029	0.592	0.108	0.017	0.017	0.017
2	0.033	0.163	0.142	0.575	0.029	0.029	0.029
3	0.034	0.018	0.185	0.717	0.015	0.015	0.015
4	0.029	0.051	0.018	0.869	0.011	0.011	0.011
5	0.038	0.036	0.045	0.036	0.773	0.036	0.037
6	0.120	0.020	0.026	0.019	0.217	0.579	0.019
7	0.014	0.015	0.014	0.013	0.187	0.119	0.637

Table 6. Estimated distribution of latent classes over time for the traditional group

Time	Latent class						
points	1	2	3	4	5	6	7
1	0.246	0.128	0.089	0.144	0.044	0.135	0.214
2	0.110	0.065	0.175	0.253	0.068	0.174	0.155
3	0.070	0.052	0.138	0.346	0.087	0.191	0.117
∞	0.042	0.063	0.070	0.530	0.111	0.135	0.048

Table 7. Estimated distribution of latent classes over time for the diagnosis group

Time	Latent class						
points	1	2	3	4	5	6	7
1	0.246	0.128	0.089	0.144	0.044	0.135	0.214
2	0.087	0.044	0.191	0.296	0.114	0.116	0.151
3	0.056	0.037	0.108	0.438	0.151	0.098	0.112
∞	0.042	0.049	0.062	0.634	0.124	0.047	0.043

4 is the state that the students are transitioning into. Otherwise, the proportions of students who are in states 3, 5, 6, and 7 are all relatively similar between the two groups.

It is possible to evaluate the long-term effects of an intervention by calculating the stationary distribution from the estimated transition distribution. The stationary distribution provides information

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

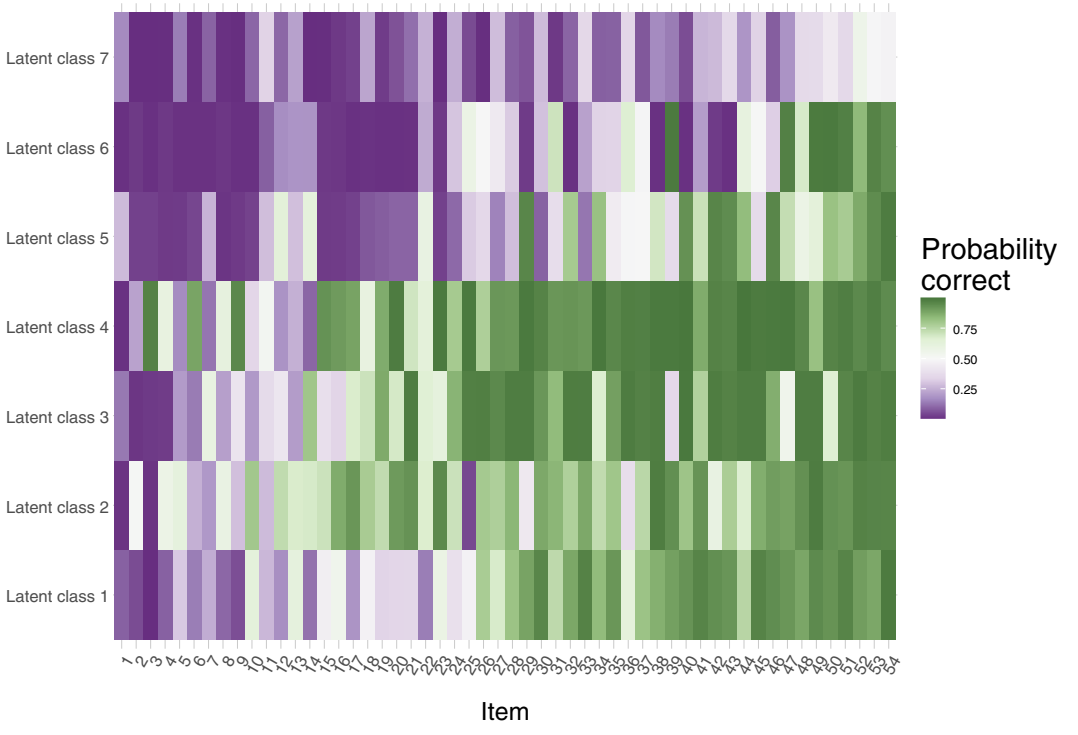


Figure 8. Probability of correct responses of all items and latent classes.

about the expected proportion of individuals in each latent class. In Tables 6 and 7, the stationary distribution is the row with time point ∞. Observe that the results corroborate the finding that students seem to be transitioning into state 4. The stationary distributions suggest some distinction between the interventions as individuals in the traditional group have reasonable probabilities to be in states 4, 5, and 6. On the other hand, compared to the traditional group, students in the diagnosis group tend to be equally likely in states 4 and 5 and less likely to be in state 6.

It should be noted that the conventional HMM here does not put restrictions on the discrete latent class space through $\boldsymbol{\Delta}$ and/or the Q-matrix. Thus, it is necessary to use the information from the estimated item parameters (i.e., the emission matrix) to create interpretations of the hidden classes. Consequently, it would then be possible to have a better understanding of the importance and usefulness of the interventions. The estimated item parameters by latent classes are shown in Figure 8, where the items have been relabeled and ordered according to the proportion of examinees who answered them correctly. State 7 seems to correspond to examinees with the least amount of mastery as they tend to answer the items incorrectly. In contrast, individuals in state 4 would report better mastery of the attributes with higher probabilities of correct response across all items. The results before suggest that students in both the diagnosis and traditional groups tend to transition into latent class 4. One may conclude that, generally, the interventions are effective at improving students' performance over time.

7. Discussion

The current article presents identifiability theory for experimental designs within the HMM framework. We focus on two common research settings: counterbalancing and multiple-group longitudinal design. Our results extend existing identifiability conditions by allowing for time-varying emission matrices

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

and group-specific transition matrices. Accordingly, these findings can provide new insights for practitioners designing assessments and studies under more practical assumptions.

The proofs and corollary in this article follow procedures illustrated by Allman et al. (2009) and Bonhomme et al. (2016). The assumptions discussed are sufficient conditions for identifiability of HMMs, but they are not necessary to the associated experimental design. Other restrictions may be imposed to identify the parameters of HMMs and RHMMs depending on the setting of interest. Future research may investigate identifiability conditions for other situations such as when there are attribute hierarchies. In addition, researchers might also want to find weaker identifiability conditions for simpler models. When the number of latent states is large, our proposed identifiability conditions for the two recommended design still hold, but there may be an issue with estimation. In such cases, one might consider more parsimonious models to realistically fit with a given sample size. In addition, researchers may be interested in establishing necessary conditions to identify RHMMs and HMMs. A key challenge is that our proofs rely on Kruskal's theorem, which only provides sufficient conditions. Different techniques such as those employed in Gu and Xu (2019) need to be extended to longitudinal settings to derive necessary identifiability conditions.

It would also be interesting to explore identifiability conditions of RHMMs for different response data types. The current work considers binary response data and binary RLCMs (in Theorem 3 and Corollary 1) because they are common in educational research. Nevertheless, the proof strategies may be extended to nominal (see Liu & Culpepper, 2025) and general-response restricted models (Lee & Gu, 2024). Another avenue for future work concerns the impact of the $\boldsymbol{Q}/\boldsymbol{\Delta}$ matrix on identifiability of RHMMs for counterbalancing design. Here, we assume that they are expert-defined, implying that we know the permutation of the latent states and can then identify the parameters. When the $\boldsymbol{Q}$ matrix is unknown, the theorems provide identifiability conditions up to label switching. For parameter estimation, Bayesian methods (e.g., MCMC algorithms) might experience within-chain label switching, which can affect the estimates. This phenomenon should be carefully examined and existing relabeling algorithms (e.g., see Erosheva & Curtis, 2017) may be extended to permute samples and address this issue.

This article also provides a Bayesian formulation for estimating parameters of HMMs in multiple-group longitudinal design. Simulation results suggest that our algorithm can recover model parameters relatively well under different settings. Nevertheless, it appears that there might be some issues with efficiently estimating some of the parameters when the number of true latent states is high. This may be because of the unstructured nature of the HMM and/or the generated items are not sufficient at separating between the latent attributes. Future research may further examine this issue and develop better estimation strategies.

The data application demonstrates that HMMs are flexible and can provide new insights into learning processes. One limitation of the application is that we fix the initial distributions of treatment groups to be the same for demonstration purpose. We again want to emphasize the importance of considering identifiability conditions while also taking into account the theoretical implications and feasibility of any given assumption. The theory and algorithm discussed allow for, but do not require, equal initial distributions between groups. This condition may hold for research studies in which individuals are randomly assigned into groups. However, in settings where the randomization occurs at the group level, the assumption of equal initial distribution might not be satisfied as one could expect variability in the distribution of latent classes across groups to reflect group-level rather than individual-level randomness. Finally, HMMs are valuable tools to understand outcomes and effects of interventions on learning trajectories. Once such outcomes are established, frameworks such as reinforcement learning can then be applied to find the best intervention strategies.

Data availability statement. Data for the empirical application is from Tang & Zhan (2021) and can be found at https://doi.org/10.3886/E153061V1. The code used to simulate data is available from the corresponding author upon request.

Funding statement. This work was partially supported by the National Science Foundation under Grant SES No. 21-50628.

Competing interests. The authors declare none.

References

Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, *37*(6A), 3099–3132.

Balamuta, J. J., & Culpepper, S. A. (2022). Exploratory restricted latent class models with monotonicity requirements under Pòlya-gamma data augmentation. *Psychometrika*, *87*(3), 903–945.

Bonhomme, S., Jochmans, K., & Robin, J.-M. (2016). Estimating multivariate latent-structure models. *The Annals of Statistics*, *44*(2), 540–563.

Bottge, B. A., Heinrichs, M., Chan, S.-Y., Mehta, Z. D., & Watson, E. (2003). Effects of video-based and applied problems on the procedural math skills of average-and low-achieving adolescents. *Journal of Special Education Technology*, *18*(2), 5–22.

Chen, Y., Culpepper, S. A., & Liang, F. (2020). A sparse latent class model for cognitive diagnosis. *Psychometrika*, *85*(1), 121–153.

Chen, Y., Culpepper, S. A., Wang, S., & Douglas, J. (2018). A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills. *Applied Psychological Measurement*, *42*(1), 5–23.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*(510), 850–866.

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*, 633–665.

Crouse, M. S., Nowak, R. D., & Baraniuk, R. G. (2002). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, *46*(4), 886–902.

Culpepper, S. A. (2023). A note on weaker conditions for identifying restricted latent class models for binary responses. *Psychometrika*, *88*(1), 158–174.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179–199.

De Leeuw, J., & Pruzansky, S. (1978). A new computational method to fit the weighted Euclidean distance model. *Psychometrika*, *43*(4), 479–490.

Erosheva, E. A., & Curtis, S. M. (2017). Dealing with reflection invariance in Bayesian factor analysis. *Psychometrika*, *82*(2), 295–307.

Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*(1), 19–40.

Gu, Y., & Xu, G. (2019). The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika*, *84*, 468–483.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210.

Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, *20*, 50–67.

Johnson, E. G. (1992). The design of the national assessment of educational progress. *Journal of Educational Measurement*, *29*(2), 95–110.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258–272.

Kaya, Y., & Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educational and Psychological Measurement*, *77*(3), 369–388.

Khatri, C., & Rao, C. R. (1968). Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, *30*(2), 167–180.

Khori, V., Shalamzari, S. A., Isanejad, A., Alizadeh, A. M., Alizadeh, S., Khodayari, S., Khodayari, H., Shahbazi, S., Zahedi, A., Sohanaki, H., Khaniki, M., Mahdian, R., Saffari, M., & Fayad, R. (2015). Effects of exercise training together with tamoxifen in reducing mammary tumor burden in mice: Possible underlying pathway of miR-21. *European Journal of Pharmacology*, *765*, 179–187.

Köhn, H.-F., Chiu, C.-Y., Oluwalana, O., Kim, H., & Wang, J. (2025). A two-step Q-matrix estimation method. *Applied Psychological Measurement*, *49*(1–2), 3–28.

Kruskal, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, *18*(2), 95–138.

Lee, S., & Gu, Y. (2024). New paradigm of identifiable general-response cognitive diagnostic models: Beyond categorical data. *Psychometrika*, *89*(4), 1304–1336.

Li, F., Cohen, A., Bottge, B., & Templin, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement*, *76*(2), 181–204.

Liu, Y., & Culpepper, S. A. (2025). Designing learning intervention studies: Identifiability of heterogeneous hidden Markov models. *Psychometrika*, *90*(4), 1258–1283.

Liu, Y., Culpepper, S. A., & Chen, Y. (2023). Identifiability of hidden Markov models for learning trajectories in cognitive diagnosis. *Psychometrika*, *88*(2), 361–386.

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Madison, M. J., & Bradshaw, L. P. (2018a). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*, 963–990.
Madison, M. J., & Bradshaw, L. P. (2018b). Evaluating intervention effects in a diagnostic classification model framework. *Journal of Educational Measurement*, *55*(1), 32–51.
Madison, M. J., Jeon, M., Cotterell, M., Haab, S., & Zor, S. (2025). TDCM: An R package for estimating longitudinal diagnostic classification models. *Multivariate Behavioral Research*, *60*(3), 518–527.
Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, *26*(2), 195–239.
Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it–Or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education*, *18*(4), 289–316.
Simola, J., Salojärvi, J., & Kojo, I. (2008). Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, *9*(4), 237–251.
Tang, F., & Zhan, P. (2021). Does diagnostic feedback promote learning? Evidence from a longitudinal cognitive diagnostic assessment. *AERA Open*, *7*, Article 23328584211060804.
van Hoorn, J., van Dijk, E., Meuwese, R., Rieffe, C., & Crone, E. A. (2016). Peer influence on prosocial behavior in adolescence. *Journal of Research on Adolescence*, *26*(1), 90–100.
Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.
von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 287–307.
Wang, S., Yang, Y., Culpepper, S. A., & Douglas, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order, hidden Markov model with covariates. *Journal of Educational and Behavioral Statistics*, *43*(1), 57–87.
Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, *45*(2), 675–707.
Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, *113*(523), 1284–1295.
Ye, S., Fellouris, G., Culpepper, S., & Douglas, J. (2016). Sequential detection of learning in cognitive diagnosis. *British Journal of Mathematical and Statistical Psychology*, *69*(2), 139–158.
Zhang, S., Wang, S., Chen, Y., & Kwon, S. (2024). *Hmcdm: Hidden Markov cognitive diagnosis models for learning*. https://github.com/tmsalab/hmcdm

Appendix

A. Terminologies and facts

In general, the proofs follow the procedure illustrated by Allman et al. (2009) and Bonhomme et al. (2016). The main idea is to view HMMs as a special case of mixture models by conditioning the observed responses onto a single common latent state. Then, the marginal distribution can be expressed as a function of the emission, transition, and initial probabilities. Moreover, the HMMs can be represented as a three-way array, and generic identifiability can be established using Kruskal's theorem for the uniqueness of three-way arrays (Kruskal, 1977). Before discussing the proof, we first introduce some terminologies and facts.

Definition A.1 (Kruskal rank). For a matrix $\mathbf{A}$, the Kruskal rank, $\text{rank}_K(\mathbf{A})$, is the largest number I such that any subset of columns of $\mathbf{A}$ has at most I linearly independent vectors.

Definition A.2 (Khatri–Rao matrix product). Let $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n) \in \mathbb{R}^{m \times n}$ and $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n) \in \mathbb{R}^{k \times n}$. The Khatri–Rao matrix product (i.e., a column-wise Kronecker product) of $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} * \mathbf{B} = (\mathbf{a}_1 \otimes \mathbf{b}_1, \ldots, \mathbf{a}_n \otimes \mathbf{b}_n),$$

where $\otimes$ denotes a Kronecker product.

We extend the notation for the Khatri–Rao product to a collection of matrices.

Definition A.3. The Khatri–Rao matrix product of n matrices $\mathbf{U}_j = (\mathbf{u}_{j,1}, \ldots, \mathbf{u}_{j,r}) \in \mathbb{R}^{M_j \times r}$ is denoted

$$\bigotimes_{j=1}^{n} \mathbf{U}_j = \left(\bigotimes_{j=1}^{n} \mathbf{u}_{j,1}, \ldots, \bigotimes_{j=1}^{n} \mathbf{u}_{j,r} \right).$$

Note that the product yields a $\left(\prod_{j=1}^{n} M_j\right) \times r$ matrix.

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Definition A.4 (Three-way array). Let $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]$ be a three-way array where each $\mathbf{T}_i$ has r columns with $\mathbf{T}$ defined as

$$\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3] = \sum_{l=1}^{r} \mathbf{t}_{1,l} \otimes \mathbf{t}_{2,l} \otimes \mathbf{t}_{3,l}, \tag{A.1}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{t}_{i,l}$ represents the l column of $\mathbf{T}_i$ for $i = 1, 2, 3$.

Theorem A.1 (Kruskal, 1977). *Consider the three-way array in Definition A.4. If*

$$rank_K(\mathbf{T}_1) + rank_K(\mathbf{T}_2) + rank_K(\mathbf{T}_3) \geq 2r + 2,$$

then $\mathbf{T}$ *uniquely determines* $\mathbf{T}_1$, $\mathbf{T}_2$, *and* $\mathbf{T}_3$ *up to label switching and rescaling of the columns.*

Lemma A.1 (Rank of matrix product). *If* $\mathbf{A}$ *is an* $m \times n$ *matrix,* $\mathbf{B}$ *is an* $n \times k$ *matrix, then* $rank(\mathbf{AB}) \leq \min(rank(\mathbf{A}), rank(\mathbf{B}))$.

B. Proof for counterbalancing (Theorem 3)

Consider (X_t, Z_t) for $t \in [T]$, $X_t \in [q_t]$, and $Z_t \in [2^K]$ where K is the number of attributes. Let there be G independent counterbalanced groups and note that independence implies that the likelihood function is $L(\theta) = \prod_{g \in [G]} L_g(\theta)$, where

$$\begin{aligned} L_g(\theta) = &\sum_{z_t \in [2^K]} \pi_{t,z_t} p(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{t-1} \mid z_t, \mathbf{B}_1, \ldots, \mathbf{B}_{t-1}, \boldsymbol{\pi}_1, \mathbf{A}) \\ &\times p(\boldsymbol{X}_t \mid z_t, \mathbf{B}_t) p(\boldsymbol{X}_{t+1}, \ldots, \boldsymbol{X}_T \mid z_t, \mathbf{B}_{t+1}, \ldots, \mathbf{B}_T, \mathbf{A}). \end{aligned} \tag{B.1}$$

Equation (B.1) corresponds to the likelihood for the gth counterbalanced group and $\mathbf{B}_t$ denote the emission matrix corresponding with the booklet administered at time t. We want to show that $L(\theta) = L(\tilde{\theta})$ if and only if $\theta = \tilde{\theta}$, where θ denotes the emission, transition, and initial probability parameters. The "if" direction is trivial. For "only if," we begin with $L(\theta) = L(\tilde{\theta})$ and show that it must imply under certain assumptions that $\theta = \tilde{\theta}$. Note that $L(\theta) = L(\tilde{\theta})$ implies that $\prod_g L_g(\theta) = \prod_g L_g(\tilde{\theta})$. Summing both sides to obtain the marginal distribution implies that $L_g(\theta) = L_g(\tilde{\theta})$ for all g. Then, for a specific group g, we can identify the parameters using techniques from Allman et al. (2009) and Bonhomme et al. (2016). The methods involve viewing HMMs as finite mixture models and then applying Kruskal's theorem for uniqueness of three-way arrays to establish identifiability.

The goal is to rewrite the likelihood function for a group g more generally as a three-way array. To do so, we introduce a lemma from Bonhomme et al. (2016).

Lemma B.1 (Bonhomme et al., 2016). *Let* X_t *be a random response conditioned on* Z_t, *satisfying the first-order Markovian assumption, with probabilities specified by the* $m \times r$ *emission matrix and stationary distribution,* $\boldsymbol{\pi} > 0$. *The emission matrix for*

1. X_t *given* Z_{t-1} *is* $\mathbf{BA}^\top$ *(backward projection);*
2. X_t *given* Z_{t+1} *is* $\mathbf{BD}_\pi \mathbf{AD}_\pi^{-1}$, *where* $\mathbf{D}_\pi = diag(\pi_1, \ldots, \pi_{2^K})$ *(forward projection).*

Based on Lemma B.1, the likelihood for a group g when $T \geq 3$ can be written as a three-way array by conditioning the responses onto a single hidden state at time $t \in (1, T)$. Consequently, Theorem 3 and Corollary 1 contain the condition of $T \geq 3$ as part of the assumption. We now consider the case with $T = 3$ to provide a more detailed examination of the identifiability proof.

When $T = 3$, Theorem 3 states that there must be three groups corresponding to different orderings of three booklets. To satisfy Part (b), assume without loss of generality that the different orderings of the booklets are

Group 1: Booklet 1	Booklet 2	Booklet 3
Group 2: Booklet 3	Booklet 1	Booklet 2
Group 3: Booklet 2	Booklet 3	Booklet 1.

Then, by conditioning on a common latent state Z_2 and depending on the order of booklets, we can turn the joint distribution of responses for examinees in Groups 1– 3 into

$$P(\boldsymbol{X} \mid Z_2) = P(\boldsymbol{X}_1 \mid Z_2) * P(\boldsymbol{X}_2 \mid Z_2) * P(\boldsymbol{X}_3 \mid Z_2). \tag{B.2}$$

The conditional probabilities can be obtained using Lemma B.1, and the marginal distribution of $\boldsymbol{X}$ can be expressed as a three-way array. Note that the three groups will have different representations since each group receives a different ordering of the booklets. Furthermore, the same booklet implies the same emission matrix, regardless of its position in the sequence. For

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

example, the emission matrix at time 2 for group 1 is identical to that at time 3 for group 2. For the remainder of this proof, we use $\mathbf{B}_i$ (instead of $\mathbf{B}_t$) to indicate the emission matrix correspond to booklet i. Let $\mathbf{T}$ be the marginal distribution of $\boldsymbol{X}$ and let $\mathbf{P}_i, \mathbf{F}_i$ denote the backward and forward projected distributions of the responses on booklet i conditioned on Z_t for $i = 1, 2, 3$. Then, the three-way array representations for the groups are

$$\text{Group 1: } \mathbf{T} = \left[\mathbf{B}_1\mathbf{D}_{\pi_1}\mathbf{A}\mathbf{D}_{\pi_1}^{-1}, \mathbf{B}_2\mathbf{D}_{\pi_2}, \mathbf{B}_3\mathbf{A}^\top\right] = \left[\mathbf{F}_1, \mathbf{B}_2\mathbf{D}_{\pi_2}, \mathbf{P}_3\right] = \sum_{l=1}^{r} \pi_l \mathbf{f}_{1,l} \otimes \mathbf{b}_{2,l} \otimes \mathbf{p}_{3,l}$$

$$\text{Group 2: } \mathbf{T} = \left[\mathbf{B}_3\mathbf{D}_{\pi_1}\mathbf{A}\mathbf{D}_{\pi_1}^{-1}, \mathbf{B}_1\mathbf{D}_{\pi_2}, \mathbf{B}_2\mathbf{A}^\top\right] = \left[\mathbf{F}_3, \mathbf{B}_1\mathbf{D}_{\pi_2}, \mathbf{P}_2\right] = \sum_{l=1}^{r} \pi_l \mathbf{f}_{3,l} \otimes \mathbf{b}_{1,l} \otimes \mathbf{p}_{2,l}$$

$$\text{Group 3: } \mathbf{T} = \left[\mathbf{B}_2\mathbf{D}_{\pi_1}\mathbf{A}\mathbf{D}_{\pi_1}^{-1}, \mathbf{B}_3\mathbf{D}_{\pi_2}, \mathbf{B}_1\mathbf{A}^\top\right] = \left[\mathbf{F}_2, \mathbf{B}_3\mathbf{D}_{\pi_2}, \mathbf{P}_1\right] = \sum_{l=1}^{r} \pi_l \mathbf{f}_{2,l} \otimes \mathbf{b}_{3,l} \otimes \mathbf{p}_{1,l},$$

where π_l is element l of the initial distribution $\boldsymbol{\pi}_1$, and $\mathbf{f}_{i,l}, \mathbf{b}_{i,l},$ and $\mathbf{p}_{i,l}$ are the l-th column of $\mathbf{F}_i, \mathbf{B}_i,$ and $\mathbf{P}_i$, respectively.

Using rank of matrix product, we can check that the $\mathbf{T}_i$'s within the three-way arrays for all groups are full column rank. Consider Group 1 as an example, we let $\mathbf{T}_1 = \mathbf{F}_1$, $\mathbf{T}_2 = \mathbf{B}_2\mathbf{D}_{\pi_2}$, and $\mathbf{T}_3 = \mathbf{P}_3$. Observe that Part (d) in Theorem 3 suggests that $rank(\mathbf{D}_{\pi_2}) = 2^K$, so $rank(\mathbf{T}_2) = rank(\mathbf{B}_2)$. Following the proof in Liu et al. (2023), Part (b) implies $rank(\mathbf{B}_2) = 2^K = rank(\mathbf{T}_2)$. Similarly, we can obtain that $\mathbf{T}_1$ and $\mathbf{T}_3$ are also full column rank. So, the condition for Kruskal's uniqueness theorem is satisfied and we can say that $\mathbf{T}_1$, $\mathbf{T}_2$, and $\mathbf{T}_3$ are unique up to a permutation of columns.

It is worth noting that the $\mathbf{T}_i$'s are identifiable only up to label switching and that the ordering of the latent classes can differ between counterbalanced groups. Nevertheless, this issue is mitigated in the context of RHMM with a known $\boldsymbol{Q}$ matrix. Specifically, the $\boldsymbol{Q}$ matrix imposes a unique labeling of the latent states which enables consistent parameter structures across the different groups. That is, if the model parameters are identified from the representation of one group, they can be considered to be identifiable for the remaining groups as well. This property is particularly important for the following proof.

Now, we need to show that the uniqueness of the $\mathbf{T}_i$'s implies the identifiability of the parameters $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \mathbf{A}, \mathbf{B}_i$ for $i \in [3]$. Due to the property of having a known $\boldsymbol{Q}$ matrix discussed above, we can consider $\mathbf{T}_2$ across all groups and get $\mathbf{B}_i\mathbf{D}_{\pi_2} = \tilde{\mathbf{B}}_i\tilde{\mathbf{D}}_{\pi_2}$. Since $\mathbf{B}_i$ are column-wise stochastic, we can left multiply both sides by $\mathbf{1}_{q_t}^\top$ to get $\mathbf{D}_{\pi_2} = \tilde{\mathbf{D}}_{\pi_2}$. Consequently, $\boldsymbol{\pi}_2 = \tilde{\boldsymbol{\pi}}_2$ and $\mathbf{B}_i = \tilde{\mathbf{B}}_i$ because $\mathbf{D}_{\pi_2}$ is full column rank with a unique inverse. Next, we examine $\mathbf{T}_3$ to obtain the expression $\mathbf{B}_i\mathbf{A}^\top = \tilde{\mathbf{B}}_1\tilde{\mathbf{A}}^\top$. We already checked that $\mathbf{B}_i = \tilde{\mathbf{B}}_i$. Since $\mathbf{B}_i$ is full column rank for $i \in [3]$, we can multiply each side of the equation by the corresponding unique Moore–Penrose inverse to obtain $\mathbf{A} = \tilde{\mathbf{A}}$. Lastly, consider $\mathbf{T}_1$ for all groups, we can simplify to get $\mathbf{D}_{\pi_1} = \tilde{\mathbf{D}}_{\pi_1}$ so $\boldsymbol{\pi}_1 = \tilde{\boldsymbol{\pi}}_1$.

Lastly, for cases where $T > 3$, similar procedures can be applied. The first step involves writing the emission matrix for observations from both the future and the past onto a common time t where all booklets appear across all groups. In this representation, the likelihood for each group can be expressed as a corresponding three-way array. The uniqueness of the emission matrices $\mathbf{B}_i$ can then be established by examining $\mathbf{T}_2$ across all groups. The identifiability of the remaining parameters can be established using $\mathbf{T}_1$ and $\mathbf{T}_3$, similar to the approach described above while using the collapsing property defined in Liu and Culpepper (2025) and the following lemma

Lemma B.2 (Khatri and Rao, 1968). *Let* $\mathbf{R}$ *and* $\mathbf{S}$ *be matrices with dimensions* $m \times p$ *and* $n \times q$*, and* $\mathbf{U}$ *and* $\mathbf{V}$ *be matrices of order* $p \times r$ *and* $q \times r$*. Then,*

$$(\mathbf{R} \otimes \mathbf{S})(\mathbf{U} * \mathbf{V}) = (\mathbf{R}\mathbf{U} * \mathbf{S}\mathbf{V}).$$

C. Proof for counterbalancing with DINA model (Corollary 1)

Liu et al. (2023) establish that assumption (b) ensures that the rank of the emission matrix is 2^K (i.e., full rank). Then, the proof is the same as in Appendix B.

Remark C.1. To satisfy Kruskal's condition, the emission matrices at two time points need to be full rank and the remaining emission matrix must have a Kruskal rank of at least 2. Achieving a Kruskal rank of at least 2 requires the columns of the emission matrix to be distinct. That is, the model needs to be able to distinguish between the different latent classes which represent various attribute profiles. For the DINA model, these restrictions impose some constraints on the $\boldsymbol{Q}_t$ matrix to ensure the model can differentiate between the 0 class and classes with only one mastered attribute. Specifically, the $\boldsymbol{Q}_t$ has to contain at least one identity matrix, which corresponds to assumption (b) of the corollary.

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

D. Proof for multiple-group longitudinal design (Theorem 4)

Consider the same setting in Appendix B, but now the emission matrices are the same across groups and we have group-specific transition matrices. Additionally, $Z_t \in [r]$ for $r \in \mathbb{N}$. We write the likelihood function as

$$L_g(\theta) = \sum_{z_t \in [r]} \pi_{t,\, z_t} p(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{t-1} \mid z_t, \mathbf{B}_1, \ldots, \mathbf{B}_{t-1}, \boldsymbol{\pi}_1, \mathbf{A}^{(g)}) \times p(\boldsymbol{X}_t \mid z_t, \mathbf{B}_t) p(\boldsymbol{X}_{t+1}, \ldots, \boldsymbol{X}_T \mid z_t, \mathbf{B}_{t+1}, \ldots, \mathbf{B}_T, \mathbf{A}^{(g)}), \tag{D.1}$$

which is the likelihood for the gth group and $\mathbf{A}^{(g)}$ denote the transition matrix for group g. Again, we consider the case with $T = 3$ and the proof for $T > 3$ uses the same extensions as the situation with the counterbalancing design. The proof is still based on Kruskal's theorem for the uniqueness of three-way arrays and techniques from Allman et al. (2009) and Bonhomme et al. (2016). Let $\mathbf{P}^{(g)}$ and $\mathbf{F}^{(g)}$ denote the backward and forward projected distributions of the responses for group g. Then, we represent the marginal distribution $\mathbf{T}$ of a group g as a three-way array:

$$\begin{aligned} \mathbf{T} &= [\mathbf{B}_1 \mathbf{D}_{\boldsymbol{\pi}_1^{(g)}} \mathbf{A}^{(g)} \mathbf{D}^{-1}_{\boldsymbol{\pi}_1^{(g)}}, \mathbf{B}_2 \mathbf{D}_{\boldsymbol{\pi}_2^{(g)}}, \mathbf{B}_3 \left(\mathbf{A}^{(g)}\right)^\top] \\ &= [\mathbf{F}^{(g)}, \mathbf{B}_2 \mathbf{D}_{\boldsymbol{\pi}_2^{(g)}}, \mathbf{P}^{(g)}] = \sum_{l=1}^{r} \pi_l^{(g)} \mathbf{f}_l^{(g)} \otimes \mathbf{b}_{2,\, l} \otimes \mathbf{p}_l^{(g)}, \end{aligned}$$

where $\pi_l^{(g)}$ is element l of the initial distribution $\boldsymbol{\pi}_1$ of group g, and $\mathbf{f}_l^{(g)}$, $\mathbf{b}_{2,\, l}$, and $\mathbf{p}_l^{(g)}$ are the l-th columns of $\mathbf{F}^{(g)}$, $\mathbf{B}_2$, and $\mathbf{P}^{(g)}$, respectively.

Using rank of matrix product, we can check that the $\mathbf{T}_i$'s within the three-way arrays for all groups are full column rank. We have $\mathbf{T}_1 = \mathbf{F}^{(g)}$, $\mathbf{T}_2 = \mathbf{B}_2 \mathbf{D}_{\boldsymbol{\pi}_2^{(g)}}$, and $\mathbf{T}_1 = \mathbf{P}^{(g)}$. Observe that Part (d) in Theorem 4 suggests that $rank(\mathbf{D}_{\boldsymbol{\pi}_2^{(g)}}) = r$, so $rank(\mathbf{T}_2) = rank(\mathbf{B}_2)$. By Part (b), $rank(\mathbf{B}_2) = r = rank(\mathbf{T}_2)$. Similarly, we can obtain that $\mathbf{T}_1$ and $\mathbf{T}_3$ are also full column rank. Then, according to Kruskal's theorem, we can say that $\mathbf{T}_1$, $\mathbf{T}_2$, and $\mathbf{T}_3$ are uniquely identified up to label switching.

Again, the $\mathbf{T}_i$'s are only identifiable up to label switching so the latent classes can differ between groups. Because the control group is assumed to stay the same over time in Theorem 4, we can use the latent states of the control group as reference to avoid this issue of label switching. Consequently, if the model parameters are identified from the representation of one group, they can be considered to be identifiable for the remaining groups.

Now, we need to show that the uniqueness of the $\mathbf{T}_i$'s implies the identifiability of the parameters $\boldsymbol{\pi}_1^{(g)}, \boldsymbol{\pi}_2^{(g)}, \mathbf{A}^{(g)}, \mathbf{B}_t$ for $t \in [3]$ for all group. First consider the control group, because the transition matrix is the identity, we have $\mathbf{T}_1 = \mathbf{B}_1$, $\mathbf{T}_2 = \mathbf{B}_2 \mathbf{D}_{\boldsymbol{\pi}_2^{(c)}}$, $\mathbf{T}_3 = \mathbf{B}_3$. It instantly follows that $\mathbf{B}_1$ and $\mathbf{B}_3$ are identified. Then, looking at $\mathbf{T}_2$, we get $\mathbf{B}_2 \mathbf{D}_{\boldsymbol{\pi}_2^{(c)}} = \tilde{\mathbf{B}}_2 \tilde{\mathbf{D}}_{\boldsymbol{\pi}_2^{(c)}}$. Since $\mathbf{B}_2$ are column-wise stochastic, we can left multiply both sides by $\mathbf{1}_{q_t}^\top$ to get $\mathbf{D}_{\boldsymbol{\pi}_2^{(c)}} = \tilde{\mathbf{D}}_{\boldsymbol{\pi}_2^{(c)}}$. Consequently, $\boldsymbol{\pi}_2^{(c)} = \tilde{\boldsymbol{\pi}}_2^{(c)}$ and $\mathbf{B}_2 = \tilde{\mathbf{B}}_2$ because $\mathbf{D}_{\boldsymbol{\pi}_2^{(c)}}$ is full column rank with a unique inverse. Moreover, $\boldsymbol{\pi}_2^{(c)} = \tilde{\boldsymbol{\pi}}_2^{(c)}$ implies that $\boldsymbol{\pi}_1^{(c)} = \tilde{\boldsymbol{\pi}}_1^{(c)}$ because $\mathbf{A}^{(c)}$ is the identity matrix.

To identify the $\boldsymbol{\pi}_1^{(g)}, \boldsymbol{\pi}_2^{(g)}, \mathbf{A}^{(g)}$ for other (treatment) groups g, we follow the procedure in Appendix B. Using $\mathbf{T}_2$, we have $\mathbf{B}_2 \mathbf{D}_{\boldsymbol{\pi}_2^{(g)}} = \tilde{\mathbf{B}}_2 \tilde{\mathbf{D}}_{\boldsymbol{\pi}_2^{(g)}}$. Since $\mathbf{B}_2$ are column-wise stochastic, we can left multiply both sides by $\mathbf{1}_{q_t}^\top$ to get $\mathbf{D}_{\boldsymbol{\pi}_2^{(g)}} = \tilde{\mathbf{D}}_{\boldsymbol{\pi}_2^{(g)}}$. Consequently, $\boldsymbol{\pi}_2^{(g)} = \tilde{\boldsymbol{\pi}}_2^{(g)}$. Next, we examine $\mathbf{T}_3$ to obtain $\mathbf{B}_3 \left(\mathbf{A}^{(g)}\right)^\top = \tilde{\mathbf{B}}_3 \left(\tilde{\mathbf{A}}^{(g)}\right)^\top$. We already checked that $\mathbf{B}_3 = \tilde{\mathbf{B}}_3$ using the control group. Since $\mathbf{B}_3$ is full column rank, we can multiply each side of the equation by the corresponding unique Moore–Penrose inverse to obtain $\mathbf{A}^{(g)} = \tilde{\mathbf{A}}^{(g)}$. Lastly, consider $\mathbf{T}_1$, we can simplify to get $\mathbf{D}_{\boldsymbol{\pi}_1^{(g)}} = \tilde{\mathbf{D}}_{\boldsymbol{\pi}_1^{(g)}}$ so $\boldsymbol{\pi}_1^{(g)} = \tilde{\boldsymbol{\pi}}_1^{(g)}$.

When $T > 3$, similar procedures can be applied. The first step involves writing the emission matrix for observations from both the future and the past onto a common time t. By applying Lemma B.2, the collapsing property (Liu & Culpepper, 2025), and the assumption for the control group, we can express the likelihood for the control group as a three-way array and identify the $\mathbf{B}_t$'s. Similarly, the likelihood for the treatment groups can also be represented by corresponding three-way arrays. With the identification of the $\mathbf{B}_t$'s, the identifiability of the remaining parameters (e.g., $\boldsymbol{\pi}_1^{(g)}, \boldsymbol{\pi}_2^{(g)}, \mathbf{A}^{(g)}$) can be established for the treatment groups, again using the collapsing property (Liu & Culpepper, 2025) together with Lemma B.2.

E. Full conditionals for Gibbs sampling

Rewrite Equation (8) to get

$$P(Z_{i,\, t} = z_{i,\, t} \mid Z_{i,\, t-1} = z_{i,\, t-1}, \mathbf{A}^{(g)}, \boldsymbol{\pi}_1^{(g)}) = \begin{cases} \prod_{z=1}^{r} \left(\pi_{1,\, z}^{(g)}\right)^{\mathbb{1}(z_{i,\, 1} = z)}, & t = 1 \\ \prod_{z'=1}^{r} \left[\prod_{z=1}^{r} \left(a_{z,\, z'}^{(g)}\right)^{\mathbb{1}(z_{i,\, t-1} = z)}\right]^{\mathbb{1}(z_{i,\, t} = z')}, & t > 1. \end{cases}$$

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

Then, Equation (6) can be expressed as

$$P(\boldsymbol{Z} \mid \mathbf{A}, \boldsymbol{\pi}_1) = \prod_{i=1}^{N}\left[\prod_{z=1}^{r}\left(\pi_{1,z}^{(g_i)}\right)^{\mathbb{1}(z_{i,1}=z)} \times \prod_{t=2}^{T}\prod_{z'=1}^{r}\prod_{z=1}^{r}\left(a_{z,z'}^{(g_i)}\right)^{\mathbb{1}(z_{i,t-1}=z)\mathbb{1}(z_{i,t}=z')}\right]. \tag{E.1}$$

For the initial distribution of a group g, we have the following full conditional distribution:

$$\begin{aligned} P(\boldsymbol{\pi}_1^{(g)} \mid \boldsymbol{Z}^{(g)}) &\propto P(\boldsymbol{Z}^{(g)} \mid \boldsymbol{\pi}_1^{(g)}) P(\boldsymbol{\pi}_1^{(g)}) \\ &\propto \prod_{z\in[r]}\left(\pi_{1,z}^{(g)}\right)^{n_{\pi_{1,z}^{(g)}}+d_{\pi_{1,z}^{(g)}}-1} \mathbb{1}(\boldsymbol{\pi}_1^{(g)} \in \mathcal{S}_r), \end{aligned}$$

where $n_{\pi_{1,z}^{(g)}} = \sum_{i=1}^{N}\mathbb{1}(z_{i,1}=z)\mathbb{1}(g_i=g)$ represents the number of individuals in group g within class z at time one. That is, the full conditional distribution of the initial distribution is Dirichlet$(n_{\pi_{1,z}^{(g)}} + d_{\pi_{1,z}^{(g)}})$. The algorithm used for simulation and data application assumes the initial distribution is the same for all group. This implies $n_{\pi_{1,z}^{(g)}} = \sum_{i=1}^{N}\mathbb{1}(z_{i,1}=z)$ which denotes the number of all individuals within class z at time one.

We next examine the full conditional distribution for the hidden states. Observe that Equation (6) and (E.1) both contain a product over the $i \in [N]$ observations. That is, the hidden states are conditionally independent over i given $\mathbf{B}, \mathbf{A},$ and $\boldsymbol{\pi}_1$. Note that $\mathbf{A}$ and $\boldsymbol{\pi}_1$ here denote the collection of transition matrices and initial distributions for all groups. So, we focus on the full conditional for the hidden states for a single individual i

$$\begin{aligned} P(\boldsymbol{Z}_i \mid \mathbf{B}, \mathbf{A}, \boldsymbol{\pi}_1, \boldsymbol{X}_i) &\propto P(\boldsymbol{X}_i \mid \mathbf{B}, \boldsymbol{Z}_i) P(\boldsymbol{Z}_i \mid \boldsymbol{\pi}_1, \mathbf{A}) \\ &= \prod_{t=1}^{T}\prod_{z=1}^{r}\left[\prod_{j\in[J]}\left(\prod_{x\in[M_j]} b_{j,x,z}^{\mathbb{1}(x_{i,t}=x)}\right)^{\mathbb{1}(j_{i,t}=j)}\right]^{\mathbb{1}(z_{i,t}=z)} \\ &\quad \times \prod_{z=1}^{r}\left(\pi_{1,z}^{(g_i)}\right)^{\mathbb{1}(z_{i,1}=z)} \\ &\quad \times \prod_{t=2}^{T}\prod_{z'=1}^{r}\left(a_{z,z'}^{(g_i)}\right)^{\mathbb{1}(z_{i,t-1}=z)\mathbb{1}(z_{i,t}=z')} \\ &= \prod_{j\in\mathbb{J}_t} b_{j,x_{i,1},z_{i,1}} \pi_{z_{i1}} \prod_{t=2}^{T} b_{j,x_{i,t},z_{i,t}} a_{z_{i,t-1},z_{i,t}}^{(g_i)}, \end{aligned}$$

where $\mathbb{J}_t$ is the set of items administered at time t. Now, the full conditional distribution for Z_{it} can be obtained in a sequential manner. First, when $t = 1$,

$$\begin{aligned} \tilde{\pi}_{i1z} &= P(Z_{i,1}=z \mid x_{i,1}, \mathbf{B}, z_{i,2}, \boldsymbol{\pi}_1, \mathbf{A}) \\ &= \frac{\left(\prod_{j\in\mathbb{J}_t} b_{j,x_{i,1},z}\right)\pi_{1,z}^{(g_i)} a_{z,z_{i,2}}^{(g_i)}}{\sum_{z=1}^{r}\left(\prod_{j\in\mathbb{J}_t} b_{j,x_{i,1},z}\right)\pi_{1,z}^{(g_i)} a_{z,z_{i,2}}^{(g_i)}}. \end{aligned} \tag{E.2}$$

For $1 < t < T$,

$$\begin{aligned} \tilde{\pi}_{itz} &= P(Z_{i,t}=z \mid x_{i,t}, \mathbf{B}, z_{i,t-1}, z_{i,t+1}, \mathbf{A}) \\ &= \frac{\left(\prod_{j\in\mathbb{J}_t} b_{j,x_{i,t},z}\right) a_{z_{i,t-1},z}^{(g_i)} a_{z,z_{i,t+1}}^{(g_i)}}{\sum_{z=1}^{r}\left(\prod_{j\in\mathbb{J}_t} b_{j,x_{i,t},z}\right) a_{z_{i,t-1},z}^{(g_i)} a_{z,z_{i,t+1}}^{(g_i)}}. \end{aligned} \tag{E.3}$$

Lastly, the conditional probability of Z_{iT}

$$\begin{aligned} \tilde{\pi}_{iTz} &= P(Z_{i,T}=z \mid x_{i,T}, \mathbf{B}, z_{i,T-1}, \mathbf{A}) \\ &= \frac{\left(\prod_{j\in\mathbb{J}_t} b_{j,x_{i,T},z}\right) a_{z_{i,T-1},z}^{(g_i)}}{\sum_{z=1}^{r}\left(\prod_{j\in\mathbb{J}_t} b_{j,x_{i,T},z}\right) a_{z_{i,T-1},z}^{(g_i)}}. \end{aligned} \tag{E.4}$$

Then, the conditional distribution for hidden states is

$$Z_{i,t} \mid x_{i,t}, \mathbf{B}, z_{i,t-1}, \mathbf{A} \sim \text{categorical}(\tilde{\boldsymbol{\pi}}_{itz}),$$

where $\tilde{\boldsymbol{\pi}}_{itz} = (\tilde{\pi}_{it1}, \ldots, \tilde{\pi}_{itr})$ for $t \in [T]$.

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press

The full conditional distribution of the columns of $\mathbf{B}$ is

$$P(\mathbf{B} \mid \boldsymbol{X}, \boldsymbol{Z}) \propto P(\boldsymbol{X} \mid \boldsymbol{Z}, \mathbf{B}) P(\mathbf{B}) \propto \prod_{z \in [r]} \prod_{j \in [J]} \prod_{x \in [M_j]} b_{j,x,z}^{n_{j,x,z} + d_{\mathbf{B}_j, z} - 1} \mathbb{1}(\boldsymbol{b}_{j,z} \in \mathcal{S}_{M_j}),$$

implying the full conditional distribution for column z of $\mathbf{B}$ is Dirichlet$(n_{j,x,z} + \boldsymbol{d}_{\mathbf{B}_j})$.

Lastly, we derive the full conditional distribution for the rows of the transition matrix. Specifically,

$$P(\mathbf{A}^{(g)} \mid \boldsymbol{Z}^{(g)}) \propto P(\boldsymbol{Z}^{(g)} \mid \mathbf{A}^{(g)}) P(\mathbf{A}^{(g)}) \propto \prod_{z \in [r]} \prod_{z' \in [r]} \left(a_{z,z'}^{(g)}\right)^{n_{A^{(g)},z,z',t} + d_{\mathbf{A}^{(g)},z'} - 1} \mathbb{1}(\boldsymbol{a}_z^{(g)} \in \mathcal{S}_r),$$

where $n_{A^{(g)},z,z',t} = \sum_{i=1}^{N} \sum_{t=2}^{T} \mathbb{1}(g_i = g) \mathbb{1}(z_{i,t-1} = z) \mathbb{1}(z_{i,t} = z')$ denotes the number of individuals in group g who goes from class z to class z' at time t. That is, the full conditional distribution for the zth row of $\mathbf{A}^{(g)}$ is $\mathbf{A}^{(g)} \mid \boldsymbol{Z}^{(g)} \sim$ Dirichlet$(n_{A^{(g)},z,z',t} + d_{\mathbf{A}^{(g)},z'})$.

Cite this article: Le, T. T., Culpepper, S. A. and Douglas, J. (2026). Longitudinal Designs for Diagnostic Models: Identification and Estimation. *Psychometrika*, 1361–1390. https://doi.org/10.1017/psy.2026.10128

https://doi.org/10.1017/psy.2026.10128 Published online by Cambridge University Press