Psychometrika (2025), **90**, 1258–1283
doi:10.1017/psy.2025.10024

THEORY AND METHODS

Designing Learning Intervention Studies: Identifiability of Heterogeneous Hidden Markov Models

Ying Liu and Steven Culpepper

Department of Statistics, University of Illinois Urbana-Champaign, Champaign, IL, USA

Corresponding author: Steven Culpepper; Email: sculpepp@illinois.edu

(Received 12 January 2025; revised 12 May 2025; accepted 22 May 2025; published online 22 July 2025)

Abstract

Hidden Markov models (HMMs) are popular for modeling complex, longitudinal data. Existing identifiability theory for conventional HMMs assume emission probabilities are constant over time and the Markov chain governing transitions among the hidden states is irreducible, which are assumptions that may not be applicable in all educational and psychological research settings. We generalize existing conditions on homogeneous HMMs by considering heterogeneous HMMs with time-varying emission probabilities and the potential for absorbing states. Researchers are investigating a family of models known as restricted HMMs (RHMMs), which combine HMMs and restricted latent class models (RLCMs) to provide fine-grained classification of educationally and psychologically relevant attribute profiles over time. These RHMMs leverage the benefits of RLCMs and HMMs to understand changes in attribute profiles within longitudinal designs. The identifiability of RHMM parameters is a critical issue for ensuring successful applications and accurate statistical inference regarding factors that impact outcomes in intervention studies. We establish identifiability conditions for RHMMs. The new identifiability conditions for heterogeneous HMMs and RHMMs provide researchers insights for designing interventions. We discuss different types of assessment designs and the implications for practice. We present an application of a heterogeneous HMM to daily measures of positive and negative affect.

Keywords: cognitive diagnosis model; heterogeneous hidden Markov models; identifiability; restricted latent class models

1. Introduction

The increasing availability of data from online learning systems, wearables, and handheld devices provides researchers with new opportunities to track development, evaluate interventions, and recommend content. These novel data structures often arrive as a multivariate time series that require methods to diagnose and classify respondent outcomes in a longitudinal fashion. Hidden Markov models (HMMs) provide a general framework for understanding changes in latent states over time. HMMs consist of two components: (1) an emission matrix $\mathbf{P}$ that describes the distribution of observed responses given latent states; and (2) a transition matrix $\mathbf{A}$ that governs the likelihood of transitioning between states over two adjacent time points. HMMs are designed to relate an observed $Y_t \in [q]$ with a hidden state $Z_t \in [r]$ for $r < q$ where we use the notation $[n] = \{1, \ldots, n\}$ to denote the set of natural numbers from 1 to $n \in \mathbb{N}$. Recent research (e.g., see Kaya & Leite, 2017; Li et al., 2016; Madison & Bradshaw, 2018a;

Madison et al., 2024; Zhan et al., 2019; Zhang & Chang, 2020) adapted the HMM framework for the diagnostic context with a class of models we refer to as restricted HMMs (RHMMs), which map Y_t onto a more parsimonious collection of binary attributes, $\boldsymbol{\alpha}_t \in \{0,1\}^K$. RHMMs combine features of cognitive diagnosis models (CDMs; e.g., see de la Torre & Douglas, 2004; Junker & Sijtsma, 2001; Roussos et al., 2007; Stout, 2002) and HMMs. Specifically, RHMMs define $\mathbf{P}$ according to a CDM to classify respondent states at a given time as well as a first-order HMM to describe changes in latent states over time.

The combination of the classical HMM framework with CDMs has the potential to advance learning research and to identify interventions that accelerate learning (Ye et al., 2016). It is therefore critical to understand technical details concerning the suitability of the HMM and RHMM frameworks for educational and psychological research. There is extensive research exploring the necessary and sufficient conditions for deploying CDMs (e.g., see Chen et al., 2015, 2020; Gu & Xu, 2021; Liu et al., 2013; Xu, 2017; Xu & Shang, 2018) and several studies are dedicated to understanding the identifiability of HMMs (Allman et al., 2009; Bonhomme et al., 2016), but there is less research available to guide applications that combine CDMs and HMMs (Liu et al., 2023). RHMMs have features of both CDMs and HMMs, so it is important to understand the assumptions of both components. Classical HMMs assume constant emission and transition probabilities over time in addition to positive long-run, stationary probabilities of residing in every state. The focus of this article is on relaxing the constant $\mathbf{P}$ and positive long-run probability assumptions, so we consider a constant transition matrix, $\mathbf{A}$, throughout the remainder of this manuscript. It is important to understand the practical implications of these two assumptions. The first assumption that $\mathbf{P}$ is constant over time implies that: (1) the support of the observed indicators, Y_t, and hidden states, Z_t, does not change over time; and (2) the observed Y_t are parallel measurements in the sense that $p(y_t \mid z_t)$ is constant across t. Accordingly, the constant $\mathbf{P}$ assumption could be reasonable whenever researchers observe a common set of measurements over time, such as settings that monitor health and well-being and public opinion with a fixed instrument. However, there are no guarantees the constant $\mathbf{P}$ assumption is feasible in all settings especially when the items change over time.

The second assumption deals with the stationary (long-run) probability of residing in the hidden states, which is denoted by $\boldsymbol{\pi} \in \{\boldsymbol{\pi} \in \mathbb{R}^r : \boldsymbol{\pi}^\top 1_r = 1,\ \pi_i > 0 \text{ for } i = 1,\ldots,r\}$ (i.e., the support for $\boldsymbol{\pi}$ is the r-dimensional simplex). The second assumption requires that $\boldsymbol{\pi} > \mathbf{0}_r$ where $\mathbf{0}_r$ is an r-vector of zeros and ">" is interpreted as an element-wise inequality. We can write $\boldsymbol{\pi}$ as a limit involving the initial distribution and the transition matrix. Specifically, the initial distribution is $\boldsymbol{\pi}_1$ where element j corresponds with $\pi_{1j} = \mathrm{P}(Z_1 = j)$. Iterating forward we find the distribution for time $t = 2$, $\boldsymbol{\pi}_2$, to have elements,

$$\mathrm{P}(Z_2 = k) = \sum_{j=1}^{r} \mathrm{P}(Z_1 = j)\mathrm{P}(Z_2 = k \mid Z_1 = j), \tag{1}$$

which is more simply stated as $\pi_{2k} = \boldsymbol{\pi}_1^\top \mathbf{A}_k$ where $\mathbf{A}_k$ is the kth column of $\mathbf{A}$. Consequently, $\boldsymbol{\pi}_2^\top = \boldsymbol{\pi}_1^\top \mathbf{A}$ and iterating forward to time n implies that the distribution of Z_n is $\boldsymbol{\pi}_n = \boldsymbol{\pi}_1^\top \mathbf{A}^{n-1}$. The long-run distribution is $\boldsymbol{\pi}^\top = \lim_{n\to\infty} \boldsymbol{\pi}_1^\top \mathbf{A}^n$. The extent to which $\boldsymbol{\pi} > \mathbf{0}_r$ is related to whether the Markov chain is irreducible. A Markov chain is irreducible if it is possible to move from a given state j to every other state k in finite time. The irreducibility assumption is violated (and $\boldsymbol{\pi}$ includes elements equal to 0) whenever $\mathbf{A}$ includes absorbing states. For instance, state k is an absorbing state if $\mathrm{P}(Z_t = j \mid Z_{t-1} = k) = 0$ for all j. The irreducibility assumption is satisfied in many content areas, but we can expect it to be violated in educational studies that advance learning. That is, interventions are designed to facilitate skill development toward the state of mastery, $\boldsymbol{\alpha} = \mathbf{1}_K$, which is an absorbing state whenever we believe it is unlikely to unlearn skills. Whereas forgetting skills may occur over a longer horizon (e.g., summer slide), the irreducibility assumption is generally untenable within shorter windows defined by typical intervention studies (e.g., days, weeks, or months).

The union of CDMs and HMMs offers a powerful framework for providing fine-grained diagnostic classifications, but care is needed to ensure foundational assumptions are consistent with application

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

domains. RHMMs have been applied to track changes in spatial rotation skills (Chen et al., 2018; Wang et al., 2018; Yigit & Douglas, 2021), introductory probability (Liu et al., 2023), middle school mathematics (Li et al., 2016), digital literacy (Liang et al., 2023), a cluster-randomized controlled trial of mathematics for students with disabilities (Madison & Bradshaw, 2018b), and geometric sequences (Chen & Culpepper, 2020). An important observation is that many applications have deployed RHMMs that relax the constant $\mathbf{P}$ and $\boldsymbol{\pi} > \mathbf{0}_r$ assumptions. For instance, prior research applied more general RHMMs that include: (1) time-varying emission probabilities with $\{\mathbf{P}_t\}_{t=1}^T$; and (2) the studies were applied in educational contexts with transition matrices that include absorbing states. Consequently, applied RHMM research has outpaced current theory related to the identifiability RHMM parameters.

RHMMs provide a useful framework for educational and psychological research. However, despite the widespread application and investigation of RHMMs, there is less research available regarding how to design intervention studies to ensure RHMM parameters are identified. Model identifiability is a critical issue for statistical inference. In the context of intervention studies, researchers need assurance that model parameters are identified in order to make claims about the effect of interventions on development.

Prior research established conditions for identifying parameters of HMMs to guide the design of diagnostic interventions. Several studies explored identifiability conditions for the case of continuous responses where $Y_t \in \mathbb{R}^d$ for $d > 0$ (Gassiat et al., 2016; Gassiat et al., 2020; Gassiat & Rousseau, 2016). We consider discrete responses, so the identifiability conditions derived under the assumption of continuous responses are not applicable for our case. There are several papers that explored the identifiability of unrestricted HMMs with constant emission and transition matrices and an irreducible transition matrix for discrete responses (Allman et al., 2009; Bonhomme et al., 2016; Cole, 2019; David et al., 2024; Tune et al., 2013). Specifically, Cole (2019) describes a log-likelihood profile method for establishing local identifiability of HMMs as well as a symbolic algebra strategy for establishing global identifiability of HMMs. One limitation of Cole (2019) is the fact that "...for most HMMs the exhaustive summary will be symbolically too complex for a symbolic algebra package to solve the appropriate equations" (p. 117). A paper from the theoretical statistics literature (Bonhomme et al., 2016) showed how to establish strict identifiability conditions for HMMs using results pertaining to the simultaneous diagonalization problem. Specifically, Bonhomme et al. (2016) showed that HMMs are strictly identified whenever $rank(\mathrm{P}) = rank(\mathbf{A}) = r$. Allman et al. (2009) proved that HMMs are generically identifiable, which means that the non-identifiable parameter values reside in a measure zero set of the larger parameter space, provided that researchers observe enough time points relative to the number of observed response patterns and the number of hidden states. Tune et al. (2013) consider an extension of Allman et al. (2009) by studying the identifiability of HMMs for a multiple observer model, which in the context of psychometrics corresponds with a model that includes several items at each time point. David et al. (2024) considers the identifiability of HMMs when external signals, such as observed time-varying covariates, are available as predictors of both the hidden Z_t and Y_t. David et al. (2024) show that when external signals are available that it is possible to establish conditions for identifying parameters of HMMs with time-varying emission and transition matrices. Recently, Liu et al. (2023) established identifiability conditions for RHMMs and showed that identification is closely linked with properties of the emission and transition matrices.

Existing research provides helpful guidelines for designing diagnostic intervention studies, but the studies are limited by several critical assumptions. First, the five studies (Allman et al., 2009; Bonhomme et al., 2016; Cole, 2019; Liu et al., 2023; Tune et al., 2013) assume constant emission and transition matrices over time in addition to an irreducible transition process (i.e., $\boldsymbol{\pi} > \mathbf{0}_r$). In contrast, we present new conditions for identifying HMMs and RHMMs with heterogeneous $\mathbf{P}$ and we allow for the possibility of absorbing states. Although the David et al. (2024) relaxes the assumption of constant $\mathbf{P}$ and $\mathbf{A}$ matrices, their identifiability conditions assume the availability of external covariates that relate to both hidden states and observed responses. In the context of psychometrics, David et al., 2024's requirement for external signals to relate to both hidden states and observed responses is analogous to educational and psychological researchers needing test-taker covariates that contribute to differential

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

item functioning by relating to both the measurement and transition models. The novel theorems in this article provide identifiability conditions without the need for external signals. Furthermore, David et al. (2024) assumes a stationary process for the observed and hidden states. Consequently, David et al. (2024) is not directly applicable in educational contexts as it requires external signals and assumes a stationary process, which does not accommodate learning and absorbing states.

The purpose of this paper is to fill the theory–practice gap by offering new insights about the identifiability of heterogeneous HMMs and RHMMs when **P** is time-varying and **A** includes absorbing states. We therefore consider the case where the emission matrix changes with time (i.e., a time-varying model). Furthermore, we discuss conditions for cases where the Markov chain both satisfies and does not satisfy the irreducibility assumption. The remainder of this paper includes five sections. The first section discusses identifiability conditions for an unrestricted heterogeneous HMM and the second section focuses on the case of RHMMs. The third section discusses the implications of the new identifiability conditions in terms of practical considerations for assessment design. The fourth section presents a Bayesian approach for estimating the model parameters of the heterogeneous HMM, a simulation study to demonstrate parameter recovery, and an application to a dataset concerning respondents daily changes in positive and negative affect. The final section discusses the implications of the results and directions for future research. Note that we include all proofs and technical details in the appendix.

2. Unrestricted heterogeneous HMMs

We begin our discussion by focusing on the finite heterogeneous HMM framework where the emission probabilities vary over time. The first subsection presents an overview of HMMs and the second subsection delves further into the identifiability of heterogeneous HMMs with time-varying emission matrices and includes two new theorems.

2.1. Overview

Consider an irreducible, aperiodic stationary Markov chain $Z_t \in [r](t = 1,2,\ldots,T)$, with a time-invariant $r \times r$ transition matrix **A**, and a stationary distribution $\boldsymbol{\pi}$ such that $\pi_j > 0$ and $\sum_{j=1}^{r} \pi_j = 1$. We assume the the observed $Y_t \in [q_t]$ are conditionally independent given the hidden states. That is, we assume $\mathrm{P}(Y_1,\ldots,Y_T \mid Z_1,\ldots,Z_T) = \prod_{t=1}^{T} \mathrm{P}(Y_t \mid Z_t)$. Let $\mathbf{P}_t$ denote the time-varying emission matrix with dimension $q_t \times r$ and $q_t \geq r$, which contains the conditional probabilities $\mathrm{P}(Y_t = y_t \mid Z_t = z_t)$, where the column entries are indexed by z_t and rows correspond to observation patterns y_t. We consider a first-order Markov model for the latent Z_t's, which assumes that given Z_{t-1}, Z_t is conditionally independent of past values for the hidden states, $\mathrm{P}(Z_t \mid Z_{t-1},\ldots,Z_2,Z_1) = \mathrm{P}(Z_t \mid Z_{t-1})$. The probability of transitioning between states over time is governed by the transition matrix **A**. That is, the probability of transitioning from state z_{t-1} at time $t-1$ to state z_t at time t is $\mathrm{P}(Z_t = z_t \mid Z_{t-1} = z_{t-1})$, which corresponds with row z_{t-1} and column z_t of **A**. Figure 1 presents a graph of an HMM where the transition matrix governs the relationship among the hidden states and the observations are conditionally independent given the Z_t's.

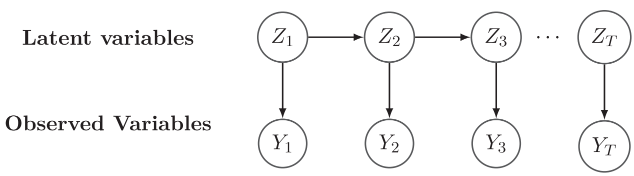

Figure 1. First-order hidden Markov model (HMM) with latent Z_t underlying the observed Y_t variables.

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

2.2. Identifiability

We next discuss the identifiability of heterogeneous HMMs. We first discuss conditions for establishing strict identifiability to ensure a unique mapping between the parameter space and likelihood function. We conclude this subsection by presenting weaker generic conditions that are needed for the non-identifiable parameter values to reside within a measure zero set.

2.2.1. Strict identifiability

Identifiability is an important prerequisite of statistical parameter inference. A model is identifiable if different values of model parameters corresponds to different probability distributions of observed variables. Model identifiability ensures that its underlying parameters can be consistently estimated.

Let $\{\mathbf{P}_t\}_{t=1}^T$ denote the T emission matrices. In the context of HMMs, we denote the parameter space of $(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ by

$$\boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) = \{(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) : \boldsymbol{\pi} \in \boldsymbol{\Omega}(\boldsymbol{\pi}), \mathbf{A} \in \boldsymbol{\Omega}(\mathbf{A}), \mathbf{P}_t \in \boldsymbol{\Omega}(\mathbf{P}_t), t = 1, 2, \ldots, T\}. \tag{2}$$

The parameter space for the stationary distribution is $\boldsymbol{\Omega}(\boldsymbol{\pi}) = \{\boldsymbol{\pi} \in [0,1]^r : \boldsymbol{\pi}^\top \mathbf{1}_r = 1\}$. The parameter spaces for $\mathbf{A}$ and $\mathbf{P}_t$ are $\boldsymbol{\Omega}(\mathbf{A}) = \{\mathbf{A} \in [0,1]^{r\times r} : \mathbf{A}\mathbf{1}_r = \mathbf{1}_r\}$ and $\boldsymbol{\Omega}(\mathbf{P}_t) = \{\mathbf{P}_t \in [0,1]^{q_t \times r} : \mathbf{P}_t^\top \mathbf{1}_{q_t} = \mathbf{1}_r\}$.

Definition 1 (Strict identifiability). The parameters $(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) \in \boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ are identifiable when

$$P(\mathbf{Y} = \mathbf{Y} \mid \boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) = P(\mathbf{Y} = \mathbf{Y} \mid \bar{\boldsymbol{\pi}}, \bar{\mathbf{A}}, \{\bar{\mathbf{P}}_t\}_{t=1}^T)$$
$$\text{if and only if } (\bar{\boldsymbol{\pi}}, \bar{\mathbf{A}}, \{\bar{\mathbf{P}}_t\}_{t=1}^T) \sim (\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T),$$

where $(\bar{\boldsymbol{\pi}}, \bar{\mathbf{A}}, \{\bar{\mathbf{P}}_t\}_{t=1}^T)$ is another value from the parameter space $\boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ and "~" means two parameter values are equivalent up to a permutation of hidden states.

The conditions shown in Assumption 1 are needed to establish the strict identifiability condition in Theorem 1.

Assumption 1. Suppose for $t \in [T]$, $Y_t \in [q_t]$ follows an HMM with hidden state $Z_t \in \{1, \ldots, r\}$, time-specific $q_t \times r$ emission matrix $\mathbf{P}_t > 0$, time-invariant transition matrix, $\mathbf{A}$, and stationary distribution with $\boldsymbol{\pi} > \mathbf{0}_r$ and $T \geq 3$ with parameters that satisfy

(a) $rank(\mathbf{A}) = r$;
(b) $\pi_c > 0$ for all $c \in [r]$;
(c) $rank_K(\mathbf{P}_{t-k}) = r$, $rank_K(\mathbf{P}_t) = r$ and $\mathbf{P}_t = \mathbf{P}_{t+1}$ for some $k \in (0,t)$ and $t \in [2, T-1]$.

Remark 1. Assumption 1b requires a positive stationary distribution and excludes the possibility of absorbing states. Assumption 1c involves a condition on the Kruskal rank of certain emission matrices. The Kruskal rank of a matrix with r columns equals r if and only if the matrix has full column rank. See Definition 13 in the Appendix A for additional details about the Kruskal rank.

We next present our Theorem related to identifiability of heterogeneous HMMs.

Theorem 1 (Strict identifiability for heterogeneous HMMs). *Under Assumption 1, then* $\mathbf{P}_1, \ldots, \mathbf{P}_T$, $\mathbf{A}$, *and* $\boldsymbol{\pi}$ *are identified, up to label-switching.*

Proof can be found in Appendix A.

Remark 2. Theorem 1 still holds if we have a slightly different condition (c) in Assumption 1 where $rank_K(\mathbf{P}_{t+k}) = r$, $rank_K(\mathbf{P}_t) = r$, and $\mathbf{P}_t = \mathbf{P}_{t-1}$ for some $k \in (t, T]$ and $t \in [2, T-1]$.

2.2.2. Generic identifiability

Theorem 1 established strict identifiability conditions for heterogeneous HMMs, which could be too strong for practical data analysis. A weaker notion of identifiability is referred to as generic identifiabilty, which was first introduced in Allman et al. (2009). Generic identifiability permits the presence of certain exceptional parameter values for which strict identifiability does not apply; however, these non-identifiable parameters must form a set with Lebesgue measure zero. Because the non-identifiable parameters are confined to a measure zero set, one is unlikely to face identifiability problems in performing inference. Therefore, generic identifiability is generally adequate for data analysis purposes. For example, Allman et al. (2009) demonstrated that generic identifiability necessitates fewer consecutive observed variables to fully determine the distribution of an HMM compared to strict identifiability. We next discuss generic identifiability of heterogeneous HMMs.

Let $S(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ denote the set of non-identifiable parameters from $\boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$:

$$
\begin{aligned}
S(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) = \{(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) :\ & P(\mathbf{Y} = \mathbf{y} \mid \boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) = P(\mathbf{Y} = \mathbf{y} \mid \bar{\boldsymbol{\pi}}, \bar{\mathbf{A}}, \{\bar{\mathbf{P}}_t\}_{t=1}^T) \text{ for} \\
& \text{some } (\bar{\boldsymbol{\pi}}, \bar{\mathbf{A}}, \{\bar{\mathbf{P}}_t\}_{t=1}^T) \nsim (\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T), (\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T) \in \boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T), \\
& (\bar{\boldsymbol{\pi}}, \bar{\mathbf{A}}, \{\bar{\mathbf{P}}_t\}_{t=1}^T) \in \boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)\}.
\end{aligned} \tag{3}
$$

Definition 2 (Generic Identifiability). The parameter space $\boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ is generically identifiable, if the Lebesgue measure of $S(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ with respect to parameter space $\boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ is zero.

Theorem 2 (Generic Identifiability for Heterogeneous HMMs). *For a heterogeneous HMM, the parameter space $\boldsymbol{\Omega}(\boldsymbol{\pi}, \mathbf{A}, \{\mathbf{P}_t\}_{t=1}^T)$ is generically identifiable up to label-switching if $\pi_c > 0$ for all $c \in [r]$, and*

(a) $\mathbf{P}_t = \mathbf{P}_{t+1}$ *for* $t \in [2, T-1]$;
(b) $\prod_{k=1}^{t-1} q_k \geq r$, $q_t \geq r$, $\prod_{k=t+1}^{T} q_k \geq r$.

Proof is shown in Appendix B. The generic condition for heterogeneous HMMs is weaker than the strict condition. Specifically, Theorem 2 imposes conditions on the dimensions of the emission matrices and does not require that any particular emission matrix is full rank.

3. Heterogeneous restricted HMMs (RHMMs)

The previous section discussed identifiability of heterogeneous HMMs where the observed response and hidden states are categorical random variables, the emission matrix $\mathbf{P}_t$ is unrestricted, and the transition matrix $\mathbf{A}$ corresponds with an irreducible and aperiodic first-order Markov process with stationary distribution $\boldsymbol{\pi}$. The purpose of this section is to introduce new identifiability conditions for restricted HMMs where restrictions are imposed upon the emission matrix. Accordingly, this section includes three subsections. The first subsection includes a discussion that connects restrictions in the emission matrix with popular diagnostic models. The second and third subsections focus on the identifiability of RHMMs for two types of models depending upon whether the transition matrix corresponds with an irreducible process. Accordingly, the second subsection corresponds with the case where $\mathbf{A}$ is irreducible with stationary distribution $\boldsymbol{\pi}$ and the third subsection presents identifiability conditions for the case where $\mathbf{A}$ includes absorbing states.

3.1. Overview and definitions

This section considers the identifiability of the heterogeneous version of RHMMs, which include a binary vector of latent attributes, $\boldsymbol{\alpha} \in \{0,1\}^K$ as opposed to the $Z \in [r]$ in the previous section. The hidden state for RHMMs at a given time t consists of a binary profile $\boldsymbol{\alpha}_t$ and indicates whether a given respondent possesses one of the 2^K possible profiles for the K attributes. We next define a general model for a categorical response $Y \in [q]$.

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

Definition 3 (Nominal diagnostic model; NDM). A general model for a categorical response $Y \in [q]$ is the NDM, which has an item response function of:

$$P(Y = m \mid \boldsymbol{\alpha}) = \frac{\exp(\mu_m(\boldsymbol{\alpha}))}{\sum_{m'=1}^{q} \exp(\mu_{m'}(\boldsymbol{\alpha}))}, \tag{4}$$

where for $m > 1$,

$$\mu_m(\boldsymbol{\alpha}) = \beta_{0m} + \sum_{k=1}^{K} \beta_{km}\alpha_k + \sum_{k<k'} \beta_{kk'm}\alpha_k\alpha_{k'} + \cdots + \beta_{12\cdots Km} \prod_{k=1}^{K} \alpha_k \tag{5}$$

includes the main-effects and interaction-effects among the attributes. Note that the parameters for the $m = 1$ case satisfies $\beta_{00} = \cdots = \beta_{12\cdots K0} = 0$ to identify the model.

Definition 4 (Structure matrix). Let $\boldsymbol{\delta}_j \in \{0,1\}^{M_j \times 2^K}$ be a binary matrix such that the element in the mth row and pth column $\delta_{jpm} = 1$ if β_{jpm} is active and non-zero and $\delta_{jpm} = 0$ if $\beta_{jpm} = 0$ and inactive.

Remark 3. The NDM is designed for nominal data and there are numerous special cases for dichotomous and polytomous response data. We refer readers to de la Torre & Douglas (2004) for a review of popular dichotomous diagnostic models. For example, Chen et al. (2020) provide an overview of how different configurations $\boldsymbol{\delta}_j$ corresponds with different diagnostic models for the $M_j = 2$ case. Additionally, there are several studies on the development of polytomous diagnostic models (e.g., see Culpepper, 2019; Fang et al., 2019; Ma & de la Torre, 2016).

An important feature for RHMMs is that we typically observe responses to multiple items at a given point in time.

Definition 5 (Multi-item emission matrix). Suppose there are J total items. We let $\boldsymbol{\theta}_j$ be the $M_j \times 2^K$ matrix of response probabilities such that element (m,c) denotes the probability of observing a response of m on item j for members of class c. We let $\mathcal{J}_t \subset [J]$ denote the subset of items administered at time t. Notice that the definition of $\mathcal{J}_t$ allows for the possibility that different items are administered over time. Accordingly, under the assumption that the $\{Y_j\}_{j\in\mathcal{J}_t}$'s are conditionally independent given $\boldsymbol{\alpha}_t$ the emission matrix for a given time t is $\mathbf{P}_t = \bigotimes_{j\in\mathcal{J}_t} \boldsymbol{\theta}_j$ where $\otimes$ is a Khatri–Rao product of matrices (see Definition 8 in Appendix A).

We next present an example to demonstrate the previously defined components of the NDM.

Example 1. Suppose $K = 2$ and $M_j = 3$, so $Y_j \in [3]$ and the $M_j \times 4$ matrix of response probabilities, $\boldsymbol{\theta}_j$, is,

$$\boldsymbol{\theta}_j = \begin{bmatrix} \theta_{j01} & \theta_{j11} & \theta_{j21} & \theta_{j31} \\ \theta_{j02} & \theta_{j12} & \theta_{j22} & \theta_{j32} \\ \theta_{j03} & \theta_{j13} & \theta_{j23} & \theta_{j33} \end{bmatrix}. \tag{6}$$

The $M_j \times 4$ matrix of regression coefficients is

$$\boldsymbol{\beta}_j = \begin{bmatrix} \beta_{j00} & \beta_{j10} & \beta_{j20} & \beta_{j30} \\ \beta_{j01} & \beta_{j11} & \beta_{j21} & \beta_{j31} \\ \beta_{j02} & \beta_{j12} & \beta_{j22} & \beta_{j32} \end{bmatrix} \tag{7}$$

and the corresponding $M_j \times 4$ structure matrix $\boldsymbol{\delta}_j$ is

$$\boldsymbol{\delta}_j = \begin{bmatrix} \delta_{j00} & \delta_{j10} & \delta_{j20} & \delta_{j30} \\ \delta_{j01} & \delta_{j11} & \delta_{j21} & \delta_{j31} \\ \delta_{j02} & \delta_{j12} & \delta_{j22} & \delta_{j32} \end{bmatrix}. \tag{8}$$

Note that $\beta_{jp0} = 0$ and $\delta_{jp0} = 0$ for all p to identify the model parameters and it is customary to include intercepts in the model, which is imposed by the restrictions $\delta_{j01} = 1$ and $\delta_{j02} = 1$.

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

3.2. Irreducible process with stationary distribution π

The purpose of this subsection is to establish the identifiability of RHMMs with an irreducible and aperiodic transition matrix $\mathbf{A}$. We first discuss conditions for strict identifiability and then conclude this subsection with results for generic identifiability.

It is important to note that there is a one-to-one mapping between the $\mathbf{P}_t$'s and the NDM θ and $\boldsymbol{\beta}$ parameters. Therefore, RHMM parameters based upon the NDM are strictly identified whenever the conditions in Assumption 1 for Theorem 1 are satisfied. However, verifying strict identifiability of the NDM can be more cumbersome given there are many parameters and ways to construct the β's so that the formed emission matrices have full column rank. Consequently, we first discuss conditions for strictly identifying RHMM parameters for the special case of binary RHMMs prior to focusing on generic identifiability for the more general setting.

Example 2 (Binary RHMMs). Suppose $M_j = 2$ for all j, so that $\mathbf{P}_t$ is a $2^{|\mathcal{J}_t|} \times 2^K$ matrix and $\boldsymbol{\beta}_j$ and $\boldsymbol{\delta}_j$ are 2^K-vectors. The structure matrix in this setting for the items administered at time t is $\boldsymbol{\Delta}_t$ is a $|\mathcal{J}_t| \times 2^K$ binary matrix with rows defined by the associated $\boldsymbol{\delta}_j$ for $j \in \mathcal{J}_t$. Accordingly, $\{\boldsymbol{\Delta}_t\}_{t=1}^T$ are the structure matrices for the items forming $\{\mathbf{P}_t\}_{t=1}^T$.

Assumption 2. Consider the conditions for the structure matrix, $\boldsymbol{\Delta}$, of binary RHMMs:

(a) $\boldsymbol{\Delta}$ takes the form $\boldsymbol{\Delta} = \left(\mathbf{D}_1^\top, \boldsymbol{\Delta}'^\top\right)^\top$ after row swapping, where $\boldsymbol{\Delta}'$ is any $(|\mathcal{J}| - K) \times 2^K$ binary matrix and $\mathbf{D}_1$ takes the form

$$\mathbb{D}_s = \left\{ D \in \{0,1\}^{K \times 2^K} : D = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & & & \vdots \\ 1 & 0 & 0 & \dots & 1 & \dots & 0 \end{bmatrix} \right\};$$

(b) For two classes of respondents, there exists at least one item in $\boldsymbol{\Delta}$, in which they have different success probabilities.

Remark 4. Note that we use the convention that the first column of $\boldsymbol{\Delta}$ corresponds to the intercept, the next K columns the main-effects for the attributes, the next $\binom{K}{2}$ columns the two-way interaction effects, etc. The last column corresponds with the K-way interaction effect.

Corollary 1 (Strict identifiability of Binary RHMMs). *Any parameter from* $\Omega_{\Delta}\left(\boldsymbol{\pi}, \mathbf{A}, \{\boldsymbol{\beta}_t\}_{t=1}^T\right)$ *is strictly identifiable up to label-switching, if:*

(1) $\mathbf{A}$ *satisfies Assumption (1a);*
(2) $\pi_c > 0$ *for all c;*
(3) $\mathbf{P}_t = \mathbf{P}_{t+1}$ *for some* $t \in [2, T-1]$*; and*
(4) $\boldsymbol{\Delta}_t$ *and* $\boldsymbol{\Delta}_{t-k}$ *satisfy Assumption (2a) for* $0 < k < t$.

Proof is shown in Appendix C.

Remark 5. Corollary 1 requires K simple structure items are administered in $\mathbf{P}_t$ and $\mathbf{P}_{t-k}$. A weaker condition is available by imposing restrictions on pairs of items as described by Culpepper (2023).

Assumption 3. Consider the conditions for the structure matrix, $\boldsymbol{\Delta}$, of binary RHMMs:

(a) $\boldsymbol{\Delta}$ takes the form $\boldsymbol{\Delta} = \left(\mathbf{D}_1^\top, \boldsymbol{\Delta}'^\top\right)^\top$ after row swapping, where $\boldsymbol{\Delta}'$ is a $(|\mathcal{J}| - K) \times 2^K$ binary matrix and $\mathbf{D}_1$ is of the following form:

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

$$\mathbb{D}_g = \left\{ D \in \{0,1\}^{K\times 2^K} : D = \begin{bmatrix} * & 1 & * & \dots & * & \dots & * \\ * & * & 1 & \dots & * & \dots & * \\ \vdots & \vdots & & \ddots & & & \vdots \\ * & * & * & \dots & 1 & \dots & * \end{bmatrix} \right\},$$

where $*$ can be either 0 or 1.

(b) In $\boldsymbol{\Delta}$, each main-effect must appear at least once. There exists a j such that $\delta_{jk} = 1$ for any $k = 1,2,\dots,K$.

Corollary 2 (Generic identifiability for binary RHMMs with $\boldsymbol{\pi} > \mathbf{0}_r$). *The parameter space $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi},\mathbf{A},\{\boldsymbol{\beta}_t\}_{t=1}^{T})$ is generically identifiable up to label-switching, if:*

(1) $\boldsymbol{\Delta}_t$ *and* $\boldsymbol{\Delta}_{t-k}$ *for* $0 < k < t$ *satisfy Assumption (3a); and*
(2) $\mathbf{P}_t = \mathbf{P}_{t+1}$ *for some* $t \in [2, T-1]$.

Proof can be found in Appendix D.

3.3. Identifiability for absorbing-state, multi-item HMM

The previously discussed classical HMM assumes that $\boldsymbol{\pi}$ is the long-run stationary distribution with the requirement that all elements are non-zero, $\boldsymbol{\pi} > \mathbf{0}_r$. The requirement of non-zero elements of $\boldsymbol{\pi}$ implies the absence of an absorbing state, which is inconsistent with the goals of education. That is, learning interventions are designed to transition students into states with greater skills, knowledge, and abilities. The purpose of this subsection is twofold. First, we present results for identifying parameters of multi-item HMMs for the case where $\mathbf{A}$ includes absorbing states. Second, we present Corollaries that are specific to the RHMM setting. We establish identifiability using Kruskal's theorem for the uniqueness of three-way arrays.

Suppose for $t \in [T]$, $Y_t \in [q_t]$ follows a multi-item HMM with hidden state $Z_t \in [r]$, time-specific $q_t \times r$ emission matrices $\{\mathbf{P}_t\}_{t=1}^{T}$, time-invariant transition matrix, $\mathbf{A}$, and an initial distribution with $\boldsymbol{\pi}_1 > 0$, two or more time points (i.e., $T > 1$) with $\mathbf{P}_1 = \mathbf{P}_{11} * \mathbf{P}_{12}$ where $\mathbf{P}_{11}$ and $\mathbf{P}_{12}$ are column-wise stochastic matrices and $*$ denotes a Khatri–Rao product, which is a column-wise Kronecker product (see Definition 8 in Appendix A for more details).

Remark 6. An important component of the absorbing state HMM is that the emission matrix for the first time point can be written as a Khatri–Rao product of two emission matrices. This assumption implies that there are two responses at time $t = 1$ with $Y_1 = (Y_{11}, Y_{12})$ such that Y_{11} and Y_{12} are conditionally independent given Z_1.

A first step for using Kruskal's theorem to establish parameters of the absorbing state RHMM are identified is to write the model as a three-way array. Specifically, we condition the observed responses on Z_1,

$$\begin{aligned} P(Y_1,\dots,Y_T \mid Z_1) &= P(Y_{11} \mid Z_1) P(Y_{12} \mid Z_1) P(Y_2,\dots,Y_T \mid Z_1) \\ &= \mathbf{P}_{11} * \mathbf{P}_{12} * \mathbf{B}_{(1,T]}, \end{aligned} \tag{9}$$

where $\mathbf{B}_{(1,T]}$ is the $(q_2 q_3 \cdots q_T) \times r$ emission matrix for $Y_2,\dots,Y_T$ given Z_1 (see Definition 9 and Equation A4 in Appendix A for more details). We write the marginal distribution, $\mathbf{M}$, of $Y_1,\dots,Y_T$ in the three-way array representation as

$$\mathbf{M} = [\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3] = [\mathbf{P}_{11}\mathbf{D}_{\pi_1}, \mathbf{P}_{12}, \mathbf{B}_{(1,T]}] = \sum_{l=1}^{r} \pi_{1l} \mathbf{p}_{11,l} \otimes \mathbf{p}_{12,l} \otimes \mathbf{b}_{(1,T],l}, \tag{10}$$

where $\mathbf{D}_{\pi_1} = \mathrm{diag}(\pi_{11},\dots,\pi_{1r})$, π_{1l} is the lth element of the initial distribution $\boldsymbol{\pi}_1$, and $\mathbf{p}_{11,l}$, $\mathbf{p}_{12,l}$, and $\mathbf{b}_{(1,T],l}$ are the lth column of $\mathbf{P}_{11}$, $\mathbf{P}_{12}$, and $\mathbf{B}_{(1,T]}$, respectively.

Theorem 3 (Strict identifiability for heterogeneous, multi-Item HMM with absorbing states). *The parameters* $\boldsymbol{\pi}_1$, $\mathbf{A}$, *and* $\{\mathbf{P}_t\}_{t=1}^T$ *of a multi-Item HMM with absorbing states are strictly identifiable up to label-switching if:*

(1) $\mathbf{A}$ *satisfies Assumption (1a);*
(2) $\pi_{1c} > 0$ *for all c;*
(3) $rank_K(\mathbf{P}_{11}) \geq 2$*,* $rank_K(\mathbf{P}_{12}) = r$*, and* $\mathbf{P}_{12} = \mathbf{P}_2$.

Proof is shown in Appendix E.

We reparameterize the multi-item HMM as a RHMM by replacing $Z_t \in [r]$ with the latent attribute profile $\boldsymbol{\alpha}_t \in \{0,1\}^K$.

Example 3 (Binary RHMMs with Absorbing States). Suppose $M_j = 2$ for all j, so that $\mathbf{P}_t$ is a $2^{|\mathcal{J}_t|} \times 2^K$ matrix and $\boldsymbol{\beta}_j$ and Δ_j are 2^K-vectors. $\boldsymbol{\Delta}_{11}$, $\boldsymbol{\Delta}_{12}$, $\{\boldsymbol{\Delta}_t\}_{t>1}$ are the structure matrices for the items forming $\mathbf{P}_{11}$, $\mathbf{P}_{12}$, and $\{\mathbf{P}_t\}_{t>1}$, respectively.

We next apply Theorem 3 to establish a corollary for the identifiability of binary RHMMs.

Corollary 3 (Strict identifiability of binary RHMMs with absorbing states). *Any parameter from* $\Omega_{\boldsymbol{\Delta}}\left(\boldsymbol{\pi}_1, \mathbf{A}, \{\boldsymbol{\beta}_t\}_{t=1}^T\right)$ *is strictly identifiable up to label-switching if:*

(1) $\mathbf{A}$ *satisfies Assumption (1a);*
(2) $\pi_{1c} > 0$ *for all c;*
(3) $\mathbf{P}_{12} = \mathbf{P}_2$*; and*
(4) $\boldsymbol{\Delta}_{12}$ *satisfies Assumption (2a), and* $\boldsymbol{\Delta}_{11}$ *satisfies Assumption (2b).*

Proof can be found in Appendix F.

Remark 7. Corollary 3 relies upon Assumption (2a), which requires that K simple structure items are administered in $\mathbf{P}_{12}$. Note that a version of Corollary 3 is possible such that strict identifiability can be achieved with weaker conditions imposed upon on pairs of items, or dyad (e.g., see Culpepper, 2023).

Corollary 4 (Generic identifiability for binary RHMMs with absorbing states). *The parameter space* $\Omega_{\boldsymbol{\Delta}}(\boldsymbol{\pi}, \mathbf{A}, \{\boldsymbol{\beta}_t\}_{t=1}^T)$ *is generically identifiable up to label-switching, if:*

(1) $\boldsymbol{\Delta}_{12}$ *satisfies Assumption (3a) and* $\boldsymbol{\Delta}_{11}$ *satisfies (3b); and*
(2) $\mathbf{P}_{12} = \mathbf{P}_2$ *for some* $t \in [2, T-1]$.

Proof can be found in Appendix G.

Remark 8. Corollaries 1, 2, 3, and 4 are focused on the case of binary response data, but these results can be easily extended to more general response models using existing results. For instance, we can extend Corollaries 1 and 3 to the case of nominal response RHMMs by replacing condition (4) with the conditions for the $\boldsymbol{\Delta}$ structure in Liu and Culpepper (2024 Theorem 1). Moreover, we can establish generic identifiability for a nominal response RHMM by replacing (1) in Corollaries 2 and 4 with the conditions for the Δ structure in Liu and Culpepper (2024, Theorem 2).

4. Practical considerations for assessment design

The purpose of this section is to provide a summary of the aforementioned results for practitioners who design assessments within the HMM or RHMM frameworks. Figure 2 includes a flowchart of the relevant decision points for deciding upon which results to use as a guide for designing studies.

The flowchart includes four layers of decision points. The first decision at the top of the flowchart requires practitioners to decide whether the emission matrix will be constant over time. An important

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

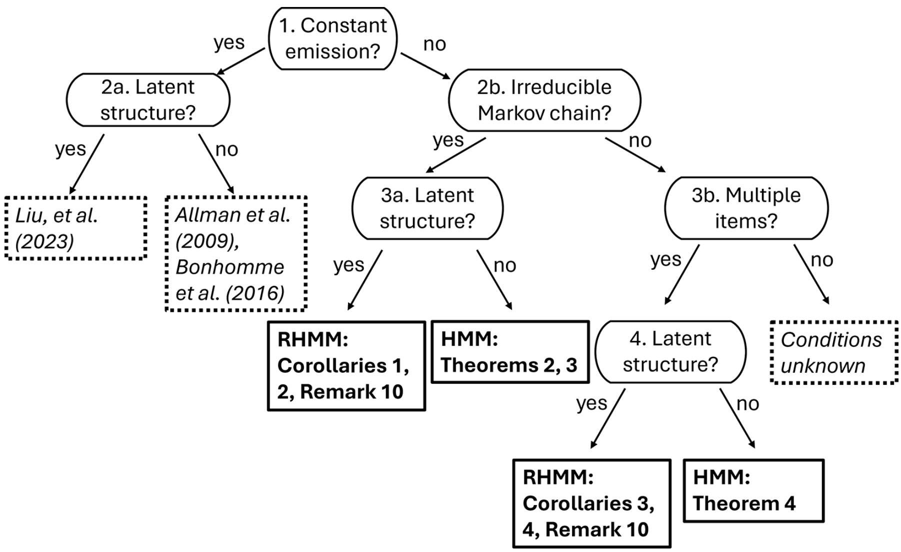

Figure 2. Flowchart of identifiability results for different types of hidden Markov models.
Note: The boxes with dotted borders and italicized text indicate conditions from existing research whereas the bold boxes indicate conditions established in this paper.

factor for determining whether $\mathbf{P}_t$ changes with time is the extent to which researchers can administer an instrument over time that has invariant response probabilities. The instrument for an invariant $\mathbf{P}$ may manifest differently for various contexts. In the case of mental health assessment, the instrument might include one or more standard items regarding patients mood or mental state. In the context of education, the instrument could consist of a collection of common items or parallel items, which require mastery of the same mental processes. If a parallel item design is deployed it is critical that the number of items and the number of response options on these items are also constant over time. In general, we should expect the emission matrices to differ in cases where different items are administered over time.

After deciding upon whether the constant emission matrix assumption is feasible researchers then move to the second layer of flowchart to (2a) if the assumption is feasible or (2b) if the emission matrix is likely to vary over time. In the case of a constant emission matrix, we direct researchers to results in previously published papers. For instance, Allman et al. (2009) and Bonhomme et al. (2016) provide results for conventional HMMs that deploy an unstructured emission matrix and Y. Liu et al. (2023) discuss conditions for identifying parameters of RHMMs. One important detail to recognize is that although Allman et al. (2009) and Bonhomme et al. (2016) describe a conventional HMM with one item per time period, their results can be applied to an emission matrix formed as a Khatri–Rao product of item emission matrices (e.g., see Definition 5).

If the answer to the constant emission matrix question at the first decision point is "no" the flowchart proceeds to question (2b), which deals with whether the Markov chain that governs transitions among hidden states is irreducible. As noted above, an irreducible Markov chain is one where it is possible to reach any state from every other state in a finite number of moves. In other words, this decision point forces researchers to grapple with the way they anticipate their phenomenon of interest to change over time. A Markov chain is irreducible if there are no absorbing states, so, researchers interested in designing assessments to study learning would likely answer "no" and continue to (3b). In other cases, researchers may expect respondents to move freely among states and it would be appropriate in these settings to proceed along the "yes" path of the flowchart to (3a).

Decision point (3a) is focused on whether the irreducible Markov chain is coupled with structure in the emission probabilities. If structure is present, the emission probabilities can be modeled as an RHMM and researchers could use binary, polytomous, or nominal restricted models for the emission response probabilities. Figure 2 indicates that the appropriate identifiability results to guide assessment design in this case are stated in Corollaries 1 and 2 and Remark 8. More specifically, RHMMs for binary data require two conditions on the emission probabilities and latent structure. First, there must be at least one pair of adjacent time points with parallel items to ensure the emission matrices are equal (e.g., condition (3) of Corollary 1 and condition (2) of Corollary 2). Second, the latent structure as described by $\boldsymbol{\Delta}$ must also satisfy conditions to ensure the minimum number of full rank emission matrices. Strict identifiability requires simple structure items as described in Assumption (2a) whereas generic identifiability requires the weaker condition presented in Assumption (3a).

If the answer of (3a) is "no," the model is a heterogeneous HMM with time-varying emission matrices. Accordingly, Theorems 1 and 2 present the appropriate conditions for designing identifiable assessment designs. Specifically, researchers that deploy unrestricted emission matrices need to administer one pair of parallel items and care is needed to ensure at least two emission matrices are full rank. One way to satisfy the parallel items assumption is to administer the same item on two adjacent time periods. Theorem 2 shows that heterogeneous HMMs are almost surely identified if the number of response options of the items over time satisfy certain inequalities (e.g., see condition (b)). For instance, the parallel item should be chosen so that the number of response options exceeds the number of hidden states. Furthermore, the product of the number of item response options administered prior to and after the parallel item must also exceed the number of classes.

Decision point (3b) focuses on whether multiple items are administered over time for the reducible Markov chain. The flowchart in Figure 2 shows that there are currently no known identifiability conditions for the case when the answer to (3b) is "no" and only a single item is administered over time.

If multiple items are administered, the flowchart directs us to our last decision point, which is whether there is structure in the response patterns of the multiple items. An answer of "no" implies the emission probabilities are unstructured, so that Theorem 3 includes the appropriate conditions for identifying the model parameters. An important feature of the conditions for the reducible Markov chain case is that conditions must be imposed upon emission probabilities from the first two time points as well as the initial distribution $\boldsymbol{\pi}_1$. In particular, the emission matrix from the first time point must consist of two booklets of items, say B_1 and B_2. In the case of heterogeneous HMMs, the model parameters will be strictly identified if the hidden states have distinct response probabilities in the emission matrix formed by B_1 and the emission matrix constructed by B_2 is full rank. Another requirement is that booklet B_2 must also be administered at the second time point in order to identify the transition matrix. Also, the initial distribution $\boldsymbol{\pi}_1$ must be strictly positive. This means that practitioners need to collect a representative sample from the population so that respondents are drawn from every hidden state. In the educational context, the assumption that $\boldsymbol{\pi}_1$ is strictly positive corresponds with ensuring the study consists of students with all types of skill profiles.

Finally, given the case at decision point (4), if the items include structure in their response probabilities the model is an RHMM and Corollaries 3 and 4 and Remark 8 provide guidelines for designing an identifiable assessment. The primary distinction between this case and the results discussed in Theorem 3 is that the Kruskal rank conditions for the emission matrices are replaced with conditions on the $\boldsymbol{\Delta}$'s. Satisfying the conditions in Assumption 2 guarantees strict identifiability for binary response RHMMs whereas the conditions in Assumption 3 are needed to generically, or almost surely, identify the binary RHMM parameters. Remark 8 discusses how previous research on the identifiability of more general nominal models can be integrated into our framework to understand the required conditions for the $\boldsymbol{\Delta}$'s.

5. Methods and empirical analyses

This section discusses issues related to the estimation and application heterogeneous HMMs. The first subsection presents a Bayesian model and full conditional distributions for a Gibbs sampling algorithm. The second subsection presents results from a simulation study to demonstrate that satisfying the

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

identifiability conditions produces a consistent estimator. The final subsection presents an application with measures of positive and negative affect.

5.1. Bayesian model for heterogeneous HMM

This subsection discusses a Bayesian model for the multi-item, heterogeneous HMM. Let $\boldsymbol{\theta}_j$ denote the $m_j \times r$ emission matrix for item $j \in [J]$ with $(\boldsymbol{\theta}_j)_{m,c}$ denoting the probability of a response of m for members of latent state c. We let $\mathcal{I}_{it}$ denote the set of items administered at time j and $J_{it} = |\mathcal{I}_{it}|$ the number of administered items. The conditional likelihood function for the response of individual i to item j at time t is a categorical distribution such that

$$Y_{itj} \mid z_{it}, \boldsymbol{\theta}_j \sim \text{categorical}\left((\boldsymbol{\theta}_j)_{y_{itj}, z_{it}}\right). \tag{11}$$

We assume the random variables within the J_t-vector of responses for individual i at time t, $\mathbf{Y}_{it} \in \{0,1\}^{J_{it}}$, are conditionally independent given Z_{it}. Accordingly, the likelihood of the responses for individual i at time t given the value of the hidden state and item parameters is

$$p(\mathbf{Y}_{it} \mid z_{it}, \boldsymbol{\Theta}_t) = \prod_{j \in \mathcal{I}_{it}} (\boldsymbol{\theta}_j)_{y_{itj}, z_{it}}, \tag{12}$$

where $\boldsymbol{\Theta}_t$ denotes the emission matrices for the administered items at time t. We consider independent categorical prior distributions for the hidden states. Specifically, the prior for respondent i is:

$$Z_{it} \mid Z_{i,t-1}, \boldsymbol{\pi}_1, \mathbf{A} \sim \begin{cases} \text{categorical}(\boldsymbol{\pi}_1) & t = 1 \\ \text{categorical}((\mathbf{A})_{z_{i,t-1},1}, \ldots, (\mathbf{A})_{z_{i,t-1},r}), & t > 1 \end{cases} \tag{13}$$

where $(\mathbf{A})_{c,c'} = \mathrm{P}(Z_{it} = c' \mid Z_{i,t-1} = c)$. We consider independent Dirichlet priors for the initial distribution, the columns of the item emission matrices, and the rows of transition matrices as

$$\boldsymbol{\pi}_1 \sim \text{Dirichlet}_r(\boldsymbol{d}_0) \tag{14}$$

$$\left((\boldsymbol{\theta}_j)_{1,c}, \ldots, (\boldsymbol{\theta}_j)_{m_j,c}\right) \sim \text{Dirichlet}_{m_j}(\boldsymbol{d}_\theta) \tag{15}$$

$$((\mathbf{A})_{c,1}, \ldots, (\mathbf{A})_{c,r}) \sim \text{Dirichlet}_r(\boldsymbol{d}_A), \tag{16}$$

where $\boldsymbol{d}_0$, $\boldsymbol{d}_\theta$, and $\boldsymbol{d}_A$ are constants and $\mathbf{W} \sim \text{Dirichlet}_k$ with density function

$$p(\mathbf{w}) = \frac{1}{B(\boldsymbol{d})} \prod_{j=1}^{k} w_i^{d_i - 1}, \tag{17}$$

where $B(\boldsymbol{d})$ denotes the multivariate beta function.

We deploy a Gibbs sampling algorithm to approximate the posterior distribution. Specifically, we sequentially sample the hidden states, initial distribution, item emission matrices, and transition matrix from the associated full conditional distributions. The full conditional distribution for the hidden state for individual i at time t is a categorical distribution, $Z_{it} \mid z_{i,t-1}, z_{i,t+1}, \boldsymbol{\pi}_1, \boldsymbol{\Theta}_t, \mathbf{A} \sim \text{categorical}(\tilde{\boldsymbol{\pi}}_{it})$, where element $c \in 1, r$ of $\tilde{\boldsymbol{\pi}}_{it}$ is

$$\tilde{\pi}_{itc} = \begin{cases} \dfrac{\left(\prod_{j \in \mathcal{I}_{i2}} (\boldsymbol{\theta}_j)_{y_{i1j},c}\right) (\boldsymbol{\pi}_1)_c \, (\mathbf{A})_{c, z_{i2}}}{\sum_{c=1}^{r} \left(\prod_{j \in \mathcal{I}_{i2}} (\boldsymbol{\theta}_j)_{y_{i1j},c}\right) (\boldsymbol{\pi}_1)_c \, (\mathbf{A})_{c, z_{i2}}} & t = 1 \\ \dfrac{\left(\prod_{j \in \mathcal{I}_{it}} (\boldsymbol{\theta}_j)_{y_{itj},c}\right) (\mathbf{A})_{z_{i,t-1},c} \, (\mathbf{A})_{c, z_{i,t+1}}}{\sum_{c=1}^{r} \left(\prod_{j \in \mathcal{I}_{it}} (\boldsymbol{\theta}_j)_{y_{itj},c}\right) (\mathbf{A})_{z_{i,t-1},c} \, (\mathbf{A})_{c, z_{i,t+1}}} & 1 < t < T \\ \dfrac{\left(\prod_{j \in \mathcal{I}_{iT}} (\boldsymbol{\theta}_j)_{y_{iTj},c}\right) (\mathbf{A})_{z_{i,T-1},c}}{\sum_{c=1}^{r} \left(\prod_{j \in \mathcal{I}_{iT}} (\boldsymbol{\theta}_j)_{y_{iTj},c}\right) (\mathbf{A})_{z_{i,T-1},c}} & t = T. \end{cases} \tag{18}$$

The full conditional distributions for the initial distribution $\boldsymbol{\pi}_1$, the columns of the item emission matrices, and the rows of the transition matrix are independent Dirichlet distributions. That is,

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

$\boldsymbol{\pi}_1 \mid z_{11}, \ldots, z_{n1} \sim \text{Dirichlet}_r(\mathbf{n}_1 + \boldsymbol{d}_1)$ where the cth element of n_1 is the number of individuals residing in state c at time $t = 1$,

$$(\mathbf{n}_1)_c = \sum_{i=1}^{n} \mathbb{1}(z_{i1} = c). \tag{19}$$

The full conditional conditional distribution for the cth column of the emission matrix for item j is

$$((\boldsymbol{\theta}_j)_{1,c}, \ldots, (\boldsymbol{\theta}_j)_{m_j,c}) \mid \mathbf{z}, \mathbf{Y} \sim \text{Dirichlet}_{m_j}(\mathbf{n}_{\theta jc} + \boldsymbol{d}_\theta), \tag{20}$$

where $\mathbf{z}$ and $\mathbf{Y}$ denote the values of the hidden states and observed responses for all individuals across time points and the kth element of $\mathbf{n}_{\theta jc}$ is the number of respondents who reside in state c and select option k on item j over time,

$$(\mathbf{n}_{\theta jc})_k = \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{1}(z_{it} = c) \mathbb{1}(j \in \mathcal{I}_{it}, y_{itj} = k). \tag{21}$$

The full conditional distribution for the cth row of $\mathbf{A}$ is

$$((\mathbf{A})_{c,1}, \ldots, (\mathbf{A})_{c,r}) \mid \mathbf{z} \sim \text{Dirichlet}_r(\mathbf{n}_{Ac} + \boldsymbol{d}_A). \tag{22}$$

The c'th element of $\mathbf{n}_{Ac}$ is the number of respondents who transition from state c to state c' over time,

$$(\mathbf{n}_{Ac})_{c'} = \sum_{i=1}^{n} \sum_{t=1}^{T-1} \mathbb{1}(z_{it} = c, z_{i,t+1} = c'). \tag{23}$$

5.2. Simulation study

We conducted a simulation study to demonstrate that enforcing the identifiability conditions produces a consistent estimator for the parameters of the heterogeneous HMM. In particular, we focus our attention to the case of the multi-item, heterogeneous HMM. Theorem 3 states that the heterogeneous HMM is identified if: (1) $rank(\mathbf{A}) = r$; (2) $\pi_{1c} > 0$ for $c \in [r]$; and (3) an item administered at time $t = 1$ has a full rank emission matrix and is also administered at time $t = 2$. We simulated data using $r = 3$ hidden states and $J = 4$ items administered over $T = 4$ time points. We simulated we responses for items 1 and 2 at time one, item 2 at time 2, item 3 at time three, and item 4 at time four. Item 2 is administered at times 1 and 2 in order to satisfy condition (3) of Theorem 3. A full column rank emission matrix was specified for the first item as

$$\boldsymbol{\theta}_1 = \begin{bmatrix} 0.70 & 0.05 & 0.05 \\ 0.10 & 0.10 & 0.05 \\ 0.10 & 0.70 & 0.10 \\ 0.05 & 0.10 & 0.10 \\ 0.05 & 0.05 & 0.70 \end{bmatrix}. \tag{24}$$

We specified distinct emission matrices for items 2 through 4 by permuting the columns of $\boldsymbol{\theta}_1$. That is, $\boldsymbol{\theta}_2$ was constructed as columns (3,2,1), $\boldsymbol{\theta}_3$ as columns (1,3,2), and $\boldsymbol{\theta}_4$ as columns (2,3,1). The data generating value for the transition matrix was

$$\mathbf{A} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}. \tag{25}$$

The data generating value for the initial distribution was $\boldsymbol{\pi}_1 = (0.375, 0.250, 0.375)$. The data generating transition matrix is full rank and the initial distribution has strictly positive elements, which implies that the data generating parameters satisfy conditions (1) and (2) of Theorem 3.

We simulated data for six conditions with samples sizes of $n = 250$, 500, 1,000, 2,000, 4,000, and 8,000. We replicated each condition 500 times and computed the mean square error (MSE) for the elements of

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

Table 1. Average MSE of the elements of $\mathbf{\Theta}$, $\mathbf{A}$, and $\boldsymbol{\pi}_1$ where MSE was computed from 500 replications

n	$\mathbf{\Theta}$	$\mathbf{A}$	$\boldsymbol{\pi}_1$
250	0.0091	0.0065	0.0029
500	0.0047	0.0035	0.0021
1,000	0.0023	0.0017	0.0014
2,000	0.0012	0.0008	0.0006
4,000	0.0007	0.0004	0.0003
8,000	0.0005	0.0003	0.0002

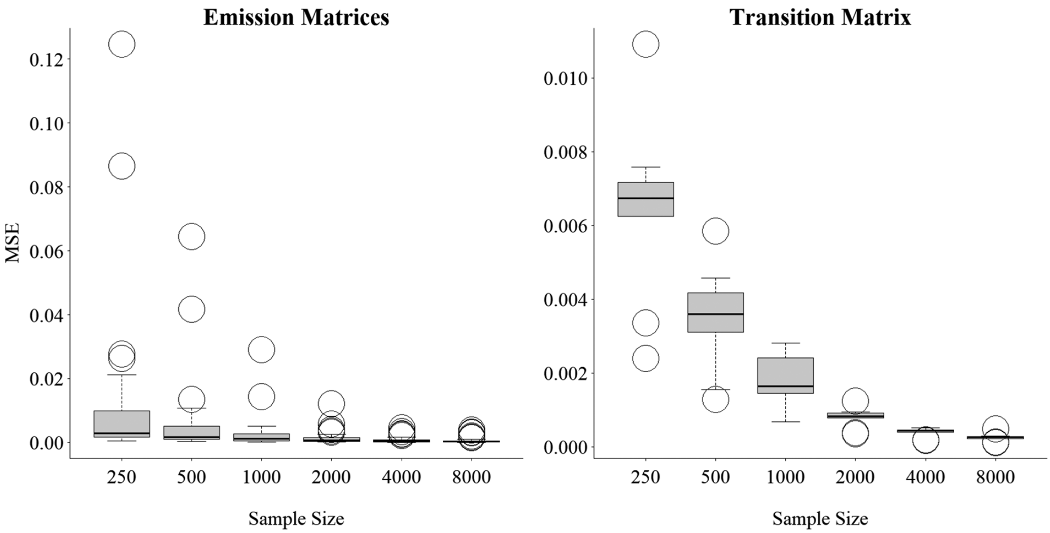

Figure 3. Boxplots of simulated MSE for elements of the emission and transition matrices by sample size.

the four emission matrices, the transition matrix, and the initial distribution. We found evidence using the Gelman-Rubin Rhat statistics of acceptable convergence using a chain length of 5,000 and a burn-in of 1,000.

Table 1 and Figure 3 show that the parameter MSE decreases as the sample size increases. Table 1 reports the average MSE for the elements of $\mathbf{\Theta}$, $\mathbf{A}$, and $\boldsymbol{\pi}_1$. The average MSE for the elements of the emission matrices equaled 0.0091 for a sample size of 250 and declined with sample size to the value of 0.0005 for $n = 8{,}000$. The average MSE for the elements of $\mathbf{A}$ and $\boldsymbol{\pi}_1$ demonstrate a similar pattern. Figure 3 provides more detailed information by presenting boxplots of the MSEs of the 60 elements of the item emission matrices and the nine element of the transition matrix by sample size. Figure 3 shows evidence of consistency in the estimation of $\mathbf{B}$ and $\mathbf{A}$ given that the MSEs decline for all elements as the sample size increases.

5.3. *Application*

Researchers are increasingly interested in developing tools to understand the dynamics of positive and negative affect, mood, and depression (Loossens et al., 2021) and recent research deployed HMMs to study changes in mood and depression (Jiang et al., 2022; Mildiner Moraga et al., 2024). Accordingly, we consider an application involving a dataset of the ten-item Positive and Negative Affect Schedule (PANAS; Shui et al., 2021). The dataset includes $n = 142$ respondents who were surveyed six times a day for five consecutive days for a maximum of 30 observations over time. Participants reported ratings on a five-point scale with anchor endpoints for the value of “1” as “not at all” to a value of “5” to indicate “extremely” for the following item stems: upset, hostile, alert, ashamed, inspired, nervous, determined, attentive, afraid, and active.

Previous research found evidence that responses to mood measures, such as the PANAS, are subject to time-of-day effects where mood changes systematically over the course of a day (Egloff et al., 1995). One implication could be that the psychometric properties of the PANAS items are a function of the time-of-day. Consequently, applying a homogeneous HMM to the PANAS response data may be too restrictive. It is therefore justified to evaluate the relative fit of a homogeneous and heterogeneous HMM to determine whether the emission matrix changes with the time of day.

An important contribution of our article is that our new identifiability theory provides the necessary insight for evaluating the extent to which emission matrices differ within a day. That is, we provide researchers with identifiability conditions for comparing of two identified versions of the homogeneous and heterogeneous HMMs. Therefore our application to the PANAS dataset compares the relative fit of a homogeneous and heterogeneous HMM.

There were instances in the PANAS data where respondents missed one of the six daily data collections. A complete dataset would include 4,260 person by time-of-day by day responses (i.e., $142 \times 6 \times 5$). The PANAS dataset included 3,789 responses, which implies that participants missed a total of 471 data collections (i.e., 11.1% of the total observations). For the purposes of demonstration, we treat the missing response data as missing at random and impute the hidden state for the missing observations within the Gibbs sampling algorithm. Specifically, the hidden states for the missing observations is imputed by sampling Z_{it} from Equation 18 by replacing the likelihood portion of the probability with the value of one.

We deployed the heterogeneous HMM with $r = 3$ by specifying a distinct emission matrix for the time-of-day (i.e., there were six different emission matrices). We used a Jeffreys’ prior for the columns of the item emission matrices and the rows of the transition matrix. Specifically, the columns of the item emission matrices were distributed as $\text{Dirichlet}_5(\mathbf{1}_5/2)$ where $\mathbf{1}_5$ is a vector of five ones and the rows of the transition matrix as $\text{Dirichlet}_3(\mathbf{1}_3/2)$. We generated starting values for the item emission matrices and initial distribution $\boldsymbol{\pi}_1$ by applying the K-means algorithm to the $3,789 \times 10$ matrix of observed items responses.

We found evidence of convergence using the Gelman-Rubin Rhat statistics by approximating the posterior distribution for both the homogeneous and heterogeneous HMMs by discarding the first 10,000 of 50,000 draws as burn-in. We computed the WAIC (Watanabe & Opper, 2010) for both models in order to evaluate relative fit of the homogeneous and heterogeneous HMMs. The WAIC was smallest for the homogeneous HMM with a value of 85813 versus 86301, which supports the conclusion that PANAS item-level emission matrices were constant over the course of the day.

Figure 4 presents posterior means of the item emission matrices as pie charts of the response probabilities for the ten items and three hidden states. Figure 4 provides evidence that members of state one report more extreme responses to every item. In contrast, the results in Figure 4 suggest that respondents within state two were more optimistic and tended to report lower levels on negative affect items. Additionally, members of state two were more likely to report being inspired, determined, attentive, and active. Finally, members in state three are similar to those in state two in terms of reporting less negative affect. However, members of state three reported lower levels of characteristics such as inspired, determined, and attentive than members of state two.

The posterior mean for the transition matrix was

$$\bar{\mathbf{A}} = \begin{bmatrix} 0.829 & 0.095 & 0.076 \\ 0.116 & 0.741 & 0.143 \\ 0.135 & 0.180 & 0.686 \end{bmatrix}. \tag{26}$$

The estimated transition matrix provides evidence that respondents were more likely to remain in their current state. For instance, respondents within the first state had an 82.9% chance of remaining in state

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

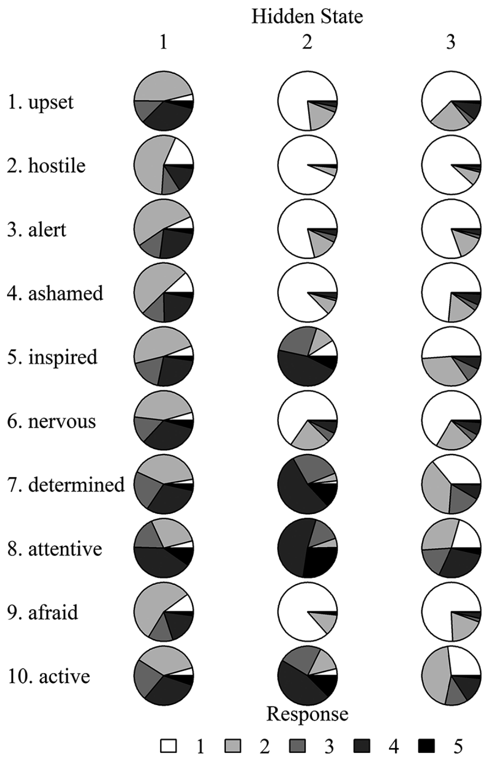

Figure 4. Pie charts of item by class emission probabilities for the PANAS dataset.
Note: The anchor labels were 1 = "not at all" and 5 = "extremely."

one and only had an 9.5% and 7.6% chance of transitioning to states two and three, respectively. In contrast, members of the optimistic state two had a 74.1% chance to remain optimistic, an 11.6% chance to transition into state one, and a 14.3% chance to transition into state three. Finally, members of state three had a 68.6% chance to remain in state three, a 13.5% chance to transition into state one, and an 18.0% chance to transition into the more optimistic state two.

6. Discussion

We presented new theory for designing assessments within the HMM and RHMM frameworks. Our results provide practitioners with new insights for creating assessment systems for monitoring student learning and changes in mental health. We extended existing identifiability theory by relaxing the constant emission matrix and irreducibility assumptions. Our results accordingly provide practitioners with guidelines for designing more flexible assessments that are grounded in assumptions that are more likely met in applied settings.

There are several directions for future research. First, we noted in the flowchart in Figure 2 that there are no known identifiability conditions for the time-varying emission matrix case when the Markov chain is reducible and a single item is administered over time. The challenge with this setting is the need to discover a technique to establish identifiability without using the forward projection properties in Lemma 2 of Appendix A derived from the stationarity assumption.

Second, we proposed a Bayesian approach for approximating parameters of the heterogeneous HMM and there are opportunities to extend the algorithm to the case of heterogeneous RHMMs. That is, researchers could consider using Bayesian variable selection methods to impose structure on the emission matrix (e.g., see Balamuta & Culpepper, 2022; Chen et al., 2020; Culpepper & Balamuta, 2023; Jimenez et al., 2023). We established identifiability conditions up to label-switching. One potential challenge with using the Bayesian framework to estimate model parameters is that MCMC algorithms might experience within-chain label-switching, which would impact parameter estimates based on posterior means. Although we did not find evidence of within-chain label-switching in our simulation study or application, within-chain label-switching should be carefully investigated. In cases where within-chain label-switching occurs, researchers should use or modify existing relabeling algorithms (e.g., see Chung, 2019; Erosheva & Curtis, 2017) to permute samples from the posterior distribution.

Third, researchers are also interested in settings where there are attribute hierarchies (e.g., see Chen & Wang, 2023; Tu et al., 2019). In the educational context, attribute hierarchies arise in cases where some attributes can only be mastered after mastery of others. For instance, suppose $K = 3$, learning is an absorbing state, and the attributes follow a linear hierarchy where students must first learn attribute 1, then attribute 2, and finally attribute 3. The flowchart in Figure 2 indicates that Corollaries 3 and 4 are the relevant conditions for this setting and a requirement is that the initial distribution $\boldsymbol{\pi}_1$ is strictly positive. However, in the case of a linear attribute structure we would expect some elements of $\boldsymbol{\pi}_1$ to be zero. For instance, probabilities involving mastery of the second and third attribute without mastery of the first would be zero (e.g., $P(\alpha_1 = 0, \alpha_2 = 1, \alpha_3 = 0)$, $P(\alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 1)$, and $P(\alpha_1 = 0, \alpha_2 = 1, \alpha_3 = 1))$ in addition to probabilities such as $P(\alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 1)$ and $P(\alpha_1 = 1, \alpha_2 = 0, \alpha_3 = 1)$. Our identifiability results can be extended to allow for attribute hierarchies. For instance, Gu & Xu (2023) studied identifiability of restricted latent class models in the presence of attribute hierarchies and their results could be incorporated to establish identifiability conditions for RHMMs when attribute hierarchies exist.

In conclusion, diagnostic models provide researchers with powerful tools for tracking changes in attributes. The increasing availability of longitudinal data will provide researchers the opportunity to leverage information to infer interventions that enhance outcomes. The results we shared in this article will provide researchers with the tools to harness the wealth information available in modern studies.

Funding statement. This research was partially supported by National Science Foundation Methodology, Measurement, and Statistics program grants 1951057 and 2150628.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

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 Pólya-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.

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

Chen, Y., Culpepper, S., 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., & Culpepper, S. A. (2020). A multivariate probit model for learning trajectories: A fine-grained evaluation of an educational intervention. *Applied Psychological Measurement*, *44*(7–8), 515–530.

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.

Chen, Y., & Wang, S. (2023). Bayesian estimation of attribute hierarchy for cognitive diagnosis models. *Journal of Educational and Behavioral Statistics*, *48*(6), 810–841.

Chung, M. (2019). A Gibbs sampling algorithm that estimates the Q-matrix for the DINA model. *Journal of Mathematical Psychology*, *93*, 102275.

Cole, D. J. (2019). Parameter redundancy and identifiability in hidden Markov models. *Metron*, *77*(2), 105–118.

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

Culpepper, S. A. (2019). An exploratory diagnostic model for ordinal responses with binary attributes: Identifiability and estimation. *Psychometrika*, *84*(4), 921–940.

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

Culpepper, S. A., & Balamuta, J. J. (2023). Inferring latent structure in polytomous data with a higher-order diagnostic model. *Multivariate Behavioral Research*, *58*(2), 368–386.

David, E., Bellot, J., Le Corff, S., & Lehéricy, L. (2024). Identifiability of discrete input–output hidden Markov models with external signals. *Statistics and Computing*, *34*(1), 54.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353.

Egloff, B., Tausch, A., Kohlmann, C.-W., & Krohne, H. W. (1995). Relationships between time of day, day of the week, and positive mood: Exploring the role of the mood measure. *Motivation and Emotion*, *19*, 99–110.

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*, 19–40.

Gassiat, É., Cleynen, A., & Robin, S. (2016). Inference in finite state space non parametric hidden Markov models and applications. *Statistics and Computing*, *26*, 61–71.

Gassiat, É., Le Corff, S., & Lehéricy, L. (2020). Identifiability and consistent estimation of nonparametric translation hidden Markov models with general state space. *Journal of Machine Learning Research*, *21*(115), 1–40.

Gassiat, É., & Rousseau, J. (2016). Nonparametric finite translation hidden Markov models and extensions. *Bernoulli*, *22*(1), 193–212. https://doi.org/10.3150/14-BEJ631

Gu, Y., & Xu, G. (2021). Sufficient and necessary conditions for the identifiability of the Q-matrix. *Statistica Sinica*, *31*, 449–472.

Gu, Y., & Xu, G. (2023). Identifiability of hierarchical latent attribute models. *Statistica Sinica*, *33*, 2561–2591.

Jiang, X., Chen, Y., Ao, N., Xiao, Y., & Du, F. (2022). A depression-risk mental pattern identified by hidden Markov model in undergraduates. *International Journal of Environmental Research and Public Health*, *19*(21), 14411.

Jimenez, A., Balamuta, J. J., & Culpepper, S. A. (2023). A sequential exploratory diagnostic model using a Pólya-gamma data augmentation strategy. *British Journal of Mathematical and Statistical Psychology*, *76*(3), 513–538.

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. G., & 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*, 167–180.

Kruskal, J. B. (1976). More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika*, *41*, 281–293.

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*, 95–138.

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.

Liang, Q., de la Torre, J., & Law, N. (2023). Latent transition cognitive diagnosis model with covariates: A three-step approach. *Journal of Educational and Behavioral Statistics*, *48*(6), 690–718.

Liu, J., Xu, G., & Ying, Z. (2013). Theory of self-learning Q-matrix. *Bernoulli*, *19*(5A), 1790–1817.

Liu, Y., & Culpepper, S. A. (2024). Restricted latent class models for nominal response data: Identifiability and estimation. *Psychometrika*, *89*, 592–625. https://doi.org/10.1007/s11336-023-09940-7

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

Loossens, T., Tuerlinckx, F., & Verdonck, S. (2021). A comparison of continuous and discrete time modeling of affective processes in terms of predictive accuracy. *Scientific Reports*, *11*(1), 6218. https://doi.org/10.1038/s41598-021-85320-4

Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 253–275.

Madison, M. J., & Bradshaw, L. (2018b). Evaluating intervention effects in a diagnostic classification model framework. *Journal of Educational Measurement*, *55*(1), 32–51.

Madison, M. J., & Bradshaw, L. P. (2018a). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*, 963–990.

Madison, M. J., Chung, S., Kim, J., & Bradshaw, L. P. (2024). Approaches to estimating longitudinal diagnostic classification models. *Behaviormetrika*, *51*(1), 7–19.

Mildiner Moraga, S., Bos, F. M., Doornbos, B., Bruggeman, R., van der Krieke, L., Snippe, E., & Aarts, E. (2024). Evidence for mood instability in patients with bipolar disorder: Applying multilevel hidden Markov modeling to intensive longitudinal ecological momentary assessment data. *Journal of Psychopathology and Clinical Science*, *133*(6), 456–468. https://doi.org/10.1037/abn0000915

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. L. (2007). The fusion model skills diagnosis system. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275–318). Cambridge University Press.

Shui, X., Zhang, M., Li, Z., Hu, X., Wang, F., & Zhang, D. (2021). A dataset of daily ambulatory psychological and physiological recording for emotion research. *Scientific Data*, *8*(1), 161.

Stirzaker, D. (2003). *Elementary probability* (2nd ed.) Cambridge University Press.

Stout, W. (2002). Psychometrics: From practice to theory and back. *Psychometrika*, *67*(4), 485–518.

Tu, D., Wang, S., Cai, Y., Douglas, J., & Chang, H.-H. (2019). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted Q-matrix design. *Applied Psychological Measurement*, *43*(4), 255–271.

Tune, P., Nguyen, H. X., & Roughan, M. (2013). Hidden Markov model identifiability via tensors. In *2013 IEEE International Symposium on Information Theory* (pp. 2299–2303).

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.

Watanabe, S., & Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(12), 3571–3594.

Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *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.

Yigit, H. D., & Douglas, J. A. (2021). First-order learning models with the GDINA: Estimation with the EM algorithm and applications. *Applied Psychological Measurement*, *45*(3), 143–158.

Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). Using JAGS for Bayesian cognitive diagnosis modeling: A tutorial. *Journal of Educational and Behavioral Statistics*, *44*(4), 473–503.

Zhang, S., & Chang, H.-H. (2020). A multilevel logistic hidden Markov model for learning under cognitive diagnosis. *Behavior Research Methods*, *52*(1), 408–421.

Appendix

A. Proof of Theorem 1

The goal of this appendix is to prove Theorem 1. To establish the result we need to understand properties of emission and transition matrices. The first subsection reviews several definitions and results pertaining to emission and transition matrices. Furthermore, we write the heterogeneous HMM as a three-way array using properties of forward- and backward-projection rules. The second subsection discusses backward and forward projections for homogeneous HMMs and then extends these results to the case of heterogeneous emission matrices. Once we are able to write the heterogeneous HMM as a three-way array we use Kruskal's theorem for the uniqueness of three-way arrays to complete the proof. Accordingly, the third shows how we can write the heterogeneous HMM as a three-way array and applies Kruskal's theorem.

A.1. Properties of emission and transition matrices

We start with by introducing some basic terminology for emission and transition matrices.

Definition 6 (Stochastic matrices). A $n \times m$ matrix $\mathbf{M}$ with non-negative entries is column-wise stochastic if $\mathbf{1}_n^\top \mathbf{M} = \mathbf{1}_m^\top$ and row-wise stochastic if $\mathbf{M}\mathbf{1}_m = \mathbf{1}_n$ where for $n \in \mathbb{N}$, $\mathbf{1}_n$ is an n-vector of ones.

Remark 9. The emission matrix $\mathbf{P}_t$ is a column-wise $q_t \times r$ stochastic matrix and the transition matrix $\mathbf{A}$ is a row-wise $r \times r$ stochastic matrix.

Lemma 1 (Properties of irreducible, aperiodic stochastic matrices). *If* $\mathbf{A}$ *is an* $r \times r$ *transition matrix for an irreducible and aperiodic Markov chain then* $\mathbf{A}^\top \boldsymbol{\pi} = \boldsymbol{\pi}$ *for* $\boldsymbol{\pi} > \mathbf{0}_r$.

Proof. See Stirzaker (2003, p. 416). □

A.2. Forward and backward projections

We follow previous strategies by writing the HMM as a special case of a mixture model using forward and backward projection rules.

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

Lemma 2 (Bonhomme et al., 2016). *Let $Y_t \in [q]$ be a random response conditioned on $Z_t \in [r]$ with probabilities specified in the $q \times r$ emission matrix $\mathbf{P}$ and stationary distribution, $\boldsymbol{\pi} > \mathbf{0}_r$. Then, the first-order Markovian assumption implies that the emission matrix for:*

1. *Y_t given Z_{t-1} is $\mathbf{P}\mathbf{A}^\top$ (backward projection); and*
2. *Y_t given Z_{t+1} is $\mathbf{P}\mathbf{D}_\pi\mathbf{A}\mathbf{D}_\pi^{-1}$ where $\mathbf{D}_\pi = diag(\pi_1, \dots, \pi_r)$ (forward projection).*

Remark 10. Note that $\tilde{\mathbf{A}} = \mathbf{D}_\pi\mathbf{A}\mathbf{D}_\pi^{-1}$ arises in the distribution for Y_t given Z_{t+1} because $\mathrm{P}(Y_t \mid Z_{t+1}) = \sum_{z=1}^{r} \mathrm{P}(Y_t \mid Z_t = z)\mathrm{P}(Z_t = z \mid Z_{t+1})$ requires the conditional distribution of Z_t given Z_{t+1}. Consequently, $\tilde{\mathbf{A}}$ is the matrix of probabilities governing transitions from Z_{t+1} to Z_t.

We next present notation that is essential for constructing the joint distribution of observed responses conditioned upon a common collection of latent states. In particular, our results make use of specialized matrix products, such as the Kronecker and Khatri–Rao products.

Definition 7. The Kronecker product two matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{p\times q}$, is denoted by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}. \tag{A1}$$

Definition 8. The Khatri–Rao matrix product (i.e., a column-wise Kronecker product) of two matrices $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_r) \in \mathbb{R}^{m\times r}$ and $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_r) \in \mathbb{R}^{n\times r}$, is denoted by

$$\mathbf{A} * \mathbf{B} = (\mathbf{a}_1 \otimes \mathbf{b}_1, \dots, \mathbf{a}_r \otimes \mathbf{b}_r). \tag{A2}$$

We present an example showing how to use the Khatri–Rao product can be used to express the emission matrix for two independent responses.

Example 4. Suppose there are two items, $Y_1 \in [q_1]$ and $Y_2 \in [q_2]$ such that the responses are conditionally independent given Z. The conditional probability for a particular response pattern for Y_1 and Y_2 is $\mathrm{P}(Y_1 = j, Y_2 = k \mid Z = z) = p_{1zj}p_{2zk}$. We let $\mathbf{p}_{iz} = (p_{iz1}, \dots, p_{iz,q_i})^\top$ denote a q_i-vector of probabilities for Y_i given $Z = z$. Also, define $\mathbf{P}_i = (\mathbf{p}_{i1}, \dots, \mathbf{p}_{ir})$ as the $q_i \times r$ matrix of response probabilities across the r classes. Therefore, the conditional independence assumption implies that the $q_1q_2 \times r$ matrix of response probabilities for Y_1 and Y_2 is $\mathbf{P}_{12} = \mathbf{P}_1 * \mathbf{P}_2$.

We present a two additional examples to demonstrate how to use the forward and backward projection rules and the Khatri–Rao product to write the emission matrix for the responses of two adjacent time points.

Example 5. Suppose Y_{T-1} and Y_T follow a first-order HMM with emission matrices $\mathbf{P}_{T-1}$ and $\mathbf{P}_T$, respectively, and transition matrix $\mathbf{A}$. Lemma 2. 1 implies that the emission matrix of Y_T given Z_{T-1} is $\mathbf{P}_T\mathbf{A}^\top$. Note that Y_{T-1} and Y_T are conditionally independent given Z_{T-1}, so the emission matrix for the combined response patterns for Y_{T-1} and Y_T given Z_{T-1} is $\mathbf{P}_{T-1} * (\mathbf{P}_T\mathbf{A}^\top)$.

Example 6. Let Y_1 and Y_2 be the first two observations of a first-order HMM with emission matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively, and transition matrix $\mathbf{A}$. Lemma 2.2 implies that the emission matrix of Y_1 given Z_2 is $\mathbf{P}_1\tilde{\mathbf{A}}$. The HMM structure implies that Y_1 and Y_2 are conditionally independent given Z_2, so the emission matrix for the combined response patterns is $\mathbf{P}_1\tilde{\mathbf{A}} * \mathbf{P}_2$.

We will apply the backward and forward projection rules to an arbitrary number of time points, so we introduce notation to represent these emission matrices.

Definition 9. Let the backward projected distribution of $Y_{t+1}, \dots, Y_T$ given Z_t be defined as $\mathbf{B}_{(t,T]}$. The distribution for $Y_{t+1}, \dots, Y_T$ given Z_{t+1} is $\mathbf{B}_{[t+1,T]}$.

Definition 10. The forward projected distribution of $Y_1, \dots, Y_{t-1}$ given Z_t is denoted as $\mathbf{F}_{[1,t)}$ and the distribution of $Y_1, \dots, Y_{t-1}$ given Z_{t-1} is denoted as $\mathbf{F}_{[1,t-1]}$.

As demonstrated by Allman et al. and Bonhomme et al., we can generalize the expressions in Examples 5 and 6 and write the emission matrix for several observations from the future or the past conditional on a common time Z_t for $1 < t < T$.

Example 7. Under the HMM structure, we can apply the backward projection rule to derive the conditional distribution of $Y_{t+1}, \dots, Y_T$ given Z_t. For instance, the conditional distribution of Y_{T-2}, Y_{T-1}, Y_T given Z_{T-2} is

$$\begin{aligned} \mathbf{B}_{[T-2,T]} &= \mathrm{P}(Y_{T-2} \mid Z_{T-2})\mathrm{P}(Y_{T-1}, Y_T \mid Z_{T-2}) = \mathbf{P}_{T-2} * \left(\mathbf{B}_{[T-1,T]}\mathbf{A}^\top\right) \\ &= \mathbf{P}_{T-2} * (\mathbf{P}_{T-1} * \mathbf{P}_T\mathbf{A}^\top)\mathbf{A}^\top. \end{aligned} \tag{A3}$$

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

Applying the identity recursively implies that the distribution of $Y_{t+1}, \ldots, Y_T$ given Z_t is,

$$\mathbf{B}_{(t,T]} = \mathbf{B}_{[t+1,T]}\mathbf{A}^\top = \left(\mathbf{P}_{t+1} * (\cdots \mathbf{P}_{T-2} * (\mathbf{P}_{T-1} * \mathbf{P}_T\mathbf{A}^\top)\mathbf{A}^\top \cdots)\right)\mathbf{A}^\top. \tag{A4}$$

Similarly, the conditional distribution of Y_1, Y_2, and Y_3 given Z_3 is

$$\mathbf{F}_{[1,3]} = \mathrm{P}(Y_1, Y_2 \mid Z_3)\mathrm{P}(Y_3 \mid Z_3) = \mathbf{F}_{[1:2]}\tilde{\mathbf{A}} * \mathbf{P}_3 = (\mathbf{P}_1\tilde{\mathbf{A}} * \mathbf{P}_2)\tilde{\mathbf{A}} * \mathbf{P}_3. \tag{A5}$$

A recursive application of the forward projection rule implies that the distribution of $Y_1, \ldots, Y_{t-1}$ given Z_t is,

$$\mathbf{F}_{[1,t)} = \mathbf{F}_{[1,t-1]}\tilde{\mathbf{A}} = \left(((\mathbf{P}_1\tilde{\mathbf{A}} * \mathbf{P}_2)\tilde{\mathbf{A}} * \mathbf{P}_3 \cdots)\tilde{\mathbf{A}} * \mathbf{P}_{t-1}\right)\tilde{\mathbf{A}}. \tag{A6}$$

The forward and backward projected emission matrices rely upon the Khatri–Rao product. The following Lemma presents identities for Khatri–Rao products that will help us to understand properties of forward and backward project emission matrices.

Lemma 3 (Khatri & Rao, 1968). *Let* $\mathbf{R}$ *and* $\mathbf{S}$ *are matrices of order* $m \times p$ *and* $n \times q$*, and* $\mathbf{U}$ *and* $\mathbf{V}$ *are matrices of order* $p \times r$ *and* $q \times r$.

(a) *Mixed product property*

$$(\mathbf{R} \otimes \mathbf{S})(\mathbf{U} * \mathbf{V}) = (\mathbf{RU}) * (\mathbf{SV}). \tag{A7}$$

(b) *Let all of the column vectors of* $\mathbf{V}$ *corresponding to independent column vectors of* $\mathbf{U}$ *be non-null. Then* $rank(\mathbf{U} * \mathbf{V}) \geq rank(\mathbf{U})$*. Similarly, if all the column vectors of* $\mathbf{U}$ *corresponding to independent column vectors of* $\mathbf{V}$ *are non-null, then* $rank(\mathbf{U} * \mathbf{V}) \geq rank(\mathbf{V})$.

The next proposition establishes several important results about forward and backward projected emission matrices.

Proposition 1 (Properties of stochastic matrices). *Consider the following properties of stochastic matrices:*

1. *If* $\mathbf{A}$ *is an* $r \times r$ *row-wise stochastic matrix that is irreducible and aperiodic then* $\mathbf{1}_r^\top \mathbf{D}_\pi \mathbf{A} \mathbf{D}_\pi^{-1} = \mathbf{1}_r^\top$.
2. *Backward and forward projection recursive identities:*
   (a) $\mathbf{B}_{(t,T]} = (\mathbf{P}_{t+1} * \mathbf{B}_{(t+1,T]})\mathbf{A}^\top$
   (b) $\mathbf{F}_{[1,t+1)} = (\mathbf{F}_{[1,t)} * \mathbf{P}_t)\tilde{\mathbf{A}}$.
3. $\mathbf{B}_{(t,T]}$ *and* $\mathbf{F}_{[1,t+1)}$ *are column-wise stochastic matrices.*

Proof. Observe that (1) is established using the property in Lemma 1 of an irreducible, aperiodic $\mathbf{A}$ that $\mathbf{A}^\top\boldsymbol{\pi} = \boldsymbol{\pi}$ and the fact that $\mathbf{1}_r^\top\mathbf{D}_\pi = \boldsymbol{\pi}^\top$. We obtain (2a) because $\mathbf{B}_{(t,T]} = \mathbf{B}_{[t+1,T]}\mathbf{A}^\top = (\mathbf{P}_{t+1} * \mathbf{B}_{(t+1,T]})\mathbf{A}^\top$ and Y_{t+1} is conditionally independent of $Y_{t+2}, \ldots, Y_T$ given Z_{t+1}. Similarly, part (2b) stems from $\mathbf{F}_{[1,t+1)} = \mathbf{F}_{[1,t]}\tilde{\mathbf{A}} = (\mathbf{F}_{[1,t)} * \mathbf{P}_t)\tilde{\mathbf{A}}$, because $Y_1, \ldots, Y_{t-1}$ and Y_t are conditionally independent given Z_t.

For (3), we note that the elements of $\mathbf{B}_{(t,T]}$ and $\mathbf{F}_{[1,t)}$ are products of non-negative elements, so all that remains to show is that the columns sum to one. We proceed with the recursion in 2a for $t < T$. Let $\mathbf{1}_{(t,T]}^\top = \mathbf{1}_{q_{t+1}}^\top \otimes \mathbf{1}_{(t+1,T]}^\top$ with $\mathbf{1}_{(T-1,T]}^\top = \mathbf{1}_{q_T}^\top$, so it is a $\left(\prod_{j=t+1}^T q_j\right)$-vector of ones. Applying Lemma 3 recursively implies $\mathbf{1}_{(t,T]}^\top\mathbf{B}_{(t,T]} = \mathbf{1}_r$. Similarly, for $t > 1$, $\mathbf{1}_{[1,t)}^\top$ is a $\left(\prod_{j=1}^{t-1} q_j\right)$-vector of ones. Recursively applying Lemma 3 with $\mathbf{1}_{[1,t)}^\top = \mathbf{1}_{[1,t-1)}^\top \otimes \mathbf{1}_{q_{t-1}}^\top$ and $\mathbf{1}_{[1,2)}^\top = \mathbf{1}_{q_1}^\top$ implies that $\mathbf{1}_{[1,t)}^\top\mathbf{F}_{[1,t)} = \mathbf{1}_r$. □

We next introduce collapsing properties of backward and forward projected emission matrices.

Definition 11. Let $\mathcal{I}_\mathcal{A} = \{x \mid x \in \mathbb{Z}, x \in \mathcal{A}\}$ denote the set of integers also in set $\mathcal{A}$, and let

$$\mathbf{1}_{\mathcal{I}_A}^\top = \begin{cases} \bigotimes_{j \in \mathcal{I}_\mathcal{A}} \mathbf{1}_{q_j}^\top & \text{if } \mathcal{I}_\mathcal{A} \neq \varnothing \\ 1 & \text{if } \mathcal{I}_\mathcal{A} = \varnothing \end{cases}.$$

Example 8. Given $\mathcal{I}_\mathcal{A} = (m, n]$, we have

$$\mathbf{1}_{\mathcal{I}_\mathcal{A}}^\top = \begin{cases} \bigotimes_{j=m+1}^n \mathbf{1}_{q_j}^\top & \text{if } m < n \\ 1 & \text{if } m \geq n \end{cases}.$$

Proposition 2 (Collapsing property). *There exists constant collapsing matrices* $\mathbf{C}_{F,[1,t),k}$ *and* $\mathbf{C}_{B,(t,T],k}$ *such that,*

(1) $\mathbf{C}_{B,(t,T],k}\mathbf{B}_{(t,T]} = \mathbf{P}_{t+k}(\mathbf{A}^\top)^k$ *for* $0 < k \leq T - t$ *and*

$$\mathbf{C}_{B,(t,T],k} = \mathbf{1}_{(t,t+k-1]}^\top \otimes \mathbf{I}_{q_{t+k}} \otimes \mathbf{1}_{(t+k,T]}^\top. \tag{A8}$$

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

(2) $\mathbf{C}_{F,[1,t),k}\mathbf{F}_{[1,t)} = \mathbf{P}_{t-k}(\tilde{\mathbf{A}})^k$ *for* $0 < k < t$ *and*

$$\mathbf{C}_{F,[1,t),k} = \mathbf{1}^\top_{[1,t-k)} \otimes \mathbf{I}_{q_{t-k}} \otimes \mathbf{1}^\top_{[t-k+1,t)}. \tag{A9}$$

Proof. We define these constant matrices recursively. For part (1), we note that the collapsing matrix can be written recursively as $\mathbf{C}_{B,(t,T],k} = \mathbf{1}^\top_{q_{t+1}} \otimes \mathbf{C}_{B,(t+1,T],k}$ for $t, t+1, \ldots, t+k-2$, where $\mathbf{C}_{B,(t+1,T],k} = \mathbf{1}^\top_{(t+1,t+k-1]} \otimes \mathbf{I}_{q_{t+k}} \otimes \mathbf{1}^\top_{(t+k,T]}$, and then $\mathbf{C}_{B,(t+k-1,T],k} = \mathbf{I}_{q_{t+k}} \otimes \mathbf{1}^\top_{(t+k,T]}$.

Applying Lemma 3 and (2a) of Proposition 1 implies the first step of the recursion yields:

$$\begin{aligned}
\mathbf{C}_{B,(t,T],k}\mathbf{B}_{(t,T]} &= \left(\mathbf{1}^\top_{q_{t+1}} \otimes \mathbf{C}_{B,(t+1,T],k}\right)\left(\mathbf{P}_{t+1} * \mathbf{B}_{(t+1,T]}\right)\mathbf{A}^\top \\
&= \left(\mathbf{1}^\top_{q_{t+1}}\mathbf{P}_{t+1} * \mathbf{C}_{B,(t+1,T],k}\mathbf{B}_{(t+1,T]}\right)\mathbf{A}^\top \\
&= \left(\mathbf{1}^\top_r * \mathbf{C}_{B,(t+1,T],k}\mathbf{B}_{(t+1,T]}\right)\mathbf{A}^\top \\
&= \mathbf{C}_{B,(t+1,T],k}\mathbf{B}_{(t+1,T]}\mathbf{A}^\top.
\end{aligned} \tag{A10}$$

Iterating forward, we can get

$$\mathbf{C}_{B,(t,T],k}\mathbf{B}_{(t,T]} = \mathbf{C}_{B,(t+k-1,T],k}\mathbf{B}_{(t+k-1,T]}\left(\mathbf{A}^\top\right)^{k-1}. \tag{A11}$$

Then using (2a) and (3) of Proposition 1, we find

$$\begin{aligned}
\mathbf{C}_{B,(t+k-1,T],k}\mathbf{B}_{(t+k-1,T]}\left(\mathbf{A}^\top\right)^{k-1} &= \left(\mathbf{I}_{q_{t+k}} \otimes \mathbf{1}^\top_{(t+k,T]}\right)\left(\mathbf{P}_{t+k} * \mathbf{B}_{(t+k,T]}\right)\mathbf{A}^\top\left(\mathbf{A}^\top\right)^{k-1} \\
&= \left(\mathbf{I}_{q_{t+k}}\mathbf{P}_{t+k} * \mathbf{1}^\top_{(t+k,T]}\mathbf{B}_{(t+k,T]}\right)\left(\mathbf{A}^\top\right)^k \\
&= \left(\mathbf{P}_{t+k} * \mathbf{1}^\top_r\right)\left(\mathbf{A}^\top\right)^k \\
&= \mathbf{P}_{t+k}\left(\mathbf{A}^\top\right)^k.
\end{aligned} \tag{A12}$$

The proof for part (2) follows in a similar fashion, but in the opposite order. That is, we let $\mathbf{C}_{F,[1,t),k} = \mathbf{C}_{F,[1,t-1),k} \otimes \mathbf{1}^\top_{q_{t-1}}$ for $t-k+2, \ldots, t-1, t$, where $\mathbf{C}_{F,[1,t-1),k} = \mathbf{1}^\top_{[1,t-k)} \otimes \mathbf{I}_{q_{t-k}} \otimes \mathbf{1}^\top_{[t-k+1,t-1)}$, and then $\mathbf{C}_{F,[1,t-k+1),k} = \mathbf{1}^\top_{[1,t-k)} \otimes \mathbf{I}_{q_{t-k}}$. The first step in the recursion is,

$$\begin{aligned}
\mathbf{C}_{F,[1,t),k}\mathbf{F}_{[1,t)} &= \left(\mathbf{C}_{F,[1,t-1),k} \otimes \mathbf{1}^\top_{q_{t-1}}\right)\left(\mathbf{F}_{[1,t-1)} * \mathbf{P}_{t-1}\right)\tilde{\mathbf{A}} \\
&= \left(\mathbf{C}_{F,[1,t-1),k}\mathbf{F}_{[1,t-1)} * \mathbf{1}^\top_{q_{t-1}}\mathbf{P}_{t-1}\right)\tilde{\mathbf{A}} \\
&= \left(\mathbf{C}_{F,[1,t-1),k}\mathbf{F}_{[1,t-1)} * \mathbf{1}^\top_r\right)\tilde{\mathbf{A}} \\
&= \mathbf{C}_{F,[1,t-1),k}\mathbf{F}_{[1,t-1)}\tilde{\mathbf{A}}.
\end{aligned} \tag{A13}$$

Iterating back we find,

$$\begin{aligned}
\mathbf{C}_{F,[1,t-k+1),k}\mathbf{F}_{[1,t-k+1)}\left(\tilde{\mathbf{A}}\right)^{k-1} &= \left(\mathbf{1}^\top_{[1,t-k)} \otimes \mathbf{I}_{q_{t-k}}\right)\left(\mathbf{F}_{[1,t-k)} * \mathbf{P}_{t-k}\right)\tilde{\mathbf{A}}\left(\tilde{\mathbf{A}}\right)^{k-1} \\
&= \left(\mathbf{1}^\top_{[1,t-k)}\mathbf{F}_{[1,t-k)} * \mathbf{I}_{q_{t-k}}\mathbf{P}_{t-k}\right)\left(\tilde{\mathbf{A}}\right)^k \\
&= \left(\mathbf{1}^\top_r * \mathbf{P}_{t-k}\right)\left(\tilde{\mathbf{A}}\right)^k \\
&= \mathbf{P}_{t-k}\left(\tilde{\mathbf{A}}\right)^k.
\end{aligned} \tag{A14}$$

□

A.3. Three-way arrays

Allman et al. (2009) proved the identifiability of time homogeneous HMMs using Kruskal's theorem for the uniqueness of three-way arrays (Kruskal, 1976, 1977). We next define a three-way array.

Definition 12 (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

$$\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3] = \sum_{l=1}^{r} \mathbf{t}_{1l} \otimes \mathbf{t}_{2l} \otimes \mathbf{t}_{3l}, \tag{A15}$$

where $\mathbf{t}_{il}$ is column l of $\mathbf{T}_i$ for $i = 1,2,3$.

The next example shows that heterogeneous HMMs with, time-varying emission probabilities can be written more generally as a three-way array.

Example 9 (Heterogeneous HMMs as a three-way array). We write the heterogeneous HMM model for $Y_1, \ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots, Y_T$ as a three-way array by conditioning on Z_t. As shown by Allman et al. (2009), the HMM conditional independence assumption and the requirement that $\boldsymbol{\pi} > \mathbf{0}_r$ for the forward-propogation rule in Lemma 2 implies that,

$$\begin{aligned} \mathrm{P}(Y_1, \ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots, Y_T \mid Z_t) &= \mathrm{P}(Y_1, \ldots, Y_{t-1} \mid Z_t)\mathrm{P}(Y_t \mid Z_t)\mathrm{P}(Y_{t+1}, \ldots, Y_T \mid Z_t) \\ &= \mathbf{F}_{[1,t)} * \mathbf{P}_t * \mathbf{B}_{(t,T]}. \end{aligned} \tag{A16}$$

We can write Equation A16 as a three-way array as shown in Definition 12. Let $\mathbf{M}$ represent the marginal distribution of $Y_1, \ldots, Y_T$ and define $\mathbf{M}_1 = \mathbf{F}_{[1,t)}$ as the emission matrix for $Y_1, \ldots, Y_{t-1}$ given Z_t, $\mathbf{M}_2 = \mathbf{P}_t\mathbf{D}_\pi$ is a rescaling of the emission matrix for Y_t given Z_t, and $\mathbf{M}_3 = \mathbf{B}_{(t,T]}$ is the emission matrix for $Y_{t+1}, \ldots, Y_T$ given Z_t. Accordingly, the three-way array representation for Equation A16 given in Definition 12 is

$$\mathbf{M} = [\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3] = [\mathbf{F}_{[1,t)}, \mathbf{P}_t\mathbf{D}_\pi, \mathbf{B}_{(t,T]}] = \sum_{l=1}^{r} \pi_l \mathbf{f}_{[1,t),l} \otimes \mathbf{p}_{t,l} \otimes \mathbf{b}_{(t,T],l}, \tag{A17}$$

where π_l is element l of the stationary distribution $\boldsymbol{\pi}$, and $\mathbf{f}_{[1,t),l}$, $\mathbf{p}_{t,l}$, and $\mathbf{b}_{(t,T],l}$ are l-th column of $\mathbf{F}_{[1,t)}$, $\mathbf{P}_t$, and $\mathbf{B}_{(t,T]}$, respectively.

Definition 13. For a matrix $\mathbf{M}$, the Kruskal rank of $\mathbf{M}$, $rank_K(\mathbf{M})$, is the largest number I such that every set of I rows in $\mathbf{M}$ are linearly independent (e.g., see Allman et al., 2009).

Theorem 4 (Kruskal, 1976, 1977). *Consider the three-way array in Definition 12. If*

$$\mathrm{rank}_K(\mathbf{T}_1) + \mathrm{rank}_K(\mathbf{T}_2) + \mathrm{rank}_K(\mathbf{T}_3) \geq 2r + 2$$

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

The proof also requires knowledge about the rank of project of matrices, so we present an existing result in the next Lemma.

Lemma 4 (Rank of matrix product). *If* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\mathbf{B} \in \mathbb{R}^{n \times k}$ *and* $rank(\mathbf{B}) = n$ *then* $rank(\mathbf{AB}) = rank(\mathbf{A})$.

The following proposition establishes the uniqueness of the three emission matrices and stationary distribution for the heterogeneous HMM in Example 9.

Proposition 3. *Consider the three-way array representation of the marginal distribution* $\mathbf{M}$ *shown in Example 9. Under Assumption 1,* $\mathbf{M}_1$, $\mathbf{M}_2$, *and* $\mathbf{M}_3$ *are uniquely identified up to label-switching.*

Proof. In Equation A17, we decompose the marginal distribution of $Y_1, \ldots, Y_T$ into three tensors $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$. According to Proposition 1, we have $\mathbf{M}_1 = \mathbf{F}_{[1,t)}$, $\mathbf{M}_2 = \mathbf{P}_t\mathbf{D}_\pi$, and $\mathbf{M}_3 = \mathbf{B}_{(t,T]}$. Note that (b) implies $rank(\mathbf{D}_\pi) = r$, so Lemma 4 implies that $rank(\mathbf{P}_t\mathbf{D}_\pi) = rank(\mathbf{P}_t)$. Therefore, (c) implies that $rank_K(\mathbf{M}_2) = r$.

Applying property (b) of Lemma 3 implies that $rank(\mathbf{P}_{t+1} * \mathbf{B}_{(t+1,T]}) \geq rank(\mathbf{P}_{t+1})$ since all columns in $\mathbf{B}_{(t+1,T]}$ are nonzero. Given $rank(\mathbf{P}_{t+1}) = r$, we can conclude that $rank_K(\mathbf{M}_3) = r$. We note that $\mathbf{M}_1 = \mathbf{F}_{[1,t)}$. We can repeatedly apply (2b) of Proposition 1 to imply that

$$\mathbf{M}_1 = \big(((\mathbf{F}_{[1,t-k)} * \mathbf{P}_{t-k})\tilde{\mathbf{A}} \cdots)\tilde{\mathbf{A}} * \mathbf{P}_{t-1}\big)\tilde{\mathbf{A}}.$$

Part (b) of Lemma 3 implies $rank_K(\mathbf{F}_{[1,t-k)} * \mathbf{P}_{t-k}) \geq rank(\mathbf{P}_{t-k})$ because the columns of $\mathbf{F}_{[1,t-k)}$ are non-null. Recall (c) implies that $rank_K(\mathbf{P}_{t-k}) = r$, so $rank_K(\mathbf{F}_{[1,t-k)} * \mathbf{P}_{t-k}) = r$. Recall that $\mathbf{F}_{[1,t-k]} = \mathbf{F}_{[1,t-k)} * \mathbf{P}_{t-k}$, so Lemma 4 implies $rank_K(\mathbf{F}_{[1,t-k]}\tilde{\mathbf{A}}) = r$, because $rank(\tilde{\mathbf{A}}) = r$. The fact that $\mathbf{F}_{[1,t-k+1)} = \mathbf{F}_{[1,t-k]}\tilde{\mathbf{A}}$ implies $rank_K(\mathbf{F}_{[1,t-k+1)}) = r$. We can continue sequentially in this fashion to show that $rank_K((\mathbf{F}_{[1,t-k+1)} * \mathbf{P}_{t-k+1})\tilde{\mathbf{A}}) = r$, $rank_K((\mathbf{F}_{[1,t-k+2)} * \mathbf{P}_{t-k+2})\tilde{\mathbf{A}}) = r$,•, $rank_K((\mathbf{F}_{[1,t-1)} * \mathbf{P}_{t-1})\tilde{\mathbf{A}}) = r$, which implies $rank_K(\mathbf{M}_1) = r$.

Now we have $rank_K(\mathbf{M}_1) + rank_K(\mathbf{M}_2) + rank_K(\mathbf{M}_3) \geq 2r + 2$, then by Theorem 4, tensors $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ are uniquely identified. □

Remark 11. Note that Proposition 3 states that the parameters are identifiable up to label-switching, but not rescaling of the columns as stated in Theorem 4. The reason for this is that $\mathbf{M}$ is a column-wise stochastic matrix, so the restriction that the columns sum to one eliminate the possibility of arbitrary rescalings. For instance, let $\mathbf{D}_s$ be a diagonal matrix with positive scalars and $\mathbf{M}_s = \mathbf{M}\mathbf{D}_s$. The requirement that $\mathbf{1}_r^\top\mathbf{M}_s = \mathbf{1}_r^\top$ for a column-wise stochastic matrix implies that $\mathbf{D}_s = \mathbf{I}_r$.

Now we are able to complete the proof for Theorem 1. Given Assumption 1, by Proposition 3, $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ are uniquely identified. Next we need to show that parameters $\boldsymbol{\pi}$, $\mathbf{A}$, and $\mathbf{P}_t$ for all t can also be identified, i.e., $\mathbf{M}_1 = \hat{\mathbf{M}}_1$, $\mathbf{M}_2 = \hat{\mathbf{M}}_2$, and $\mathbf{M}_3 = \hat{\mathbf{M}}_3$ if and only if $\mathbf{P}_t = \hat{\mathbf{P}}_t$ for all t, $\mathbf{A} = \hat{\mathbf{A}}$, and $\boldsymbol{\pi} = \hat{\boldsymbol{\pi}}$. Notice that $\mathbf{M}_2 = \hat{\mathbf{M}}_2$ implies $\mathbf{P}_t\mathbf{D}_\pi = \hat{\mathbf{P}}_t\hat{\mathbf{D}}_\pi$. Pre-multiplying by $\mathbf{1}_r^\top$ on both sides implies that $\boldsymbol{\pi} = \hat{\boldsymbol{\pi}}$ and then $\mathbf{P}_t = \hat{\mathbf{P}}_t$. Given $\mathbf{P}_t = \mathbf{P}_{t+1}$, $\mathbf{P}_{t+1} = \hat{\mathbf{P}}_{t+1}$ follows. According to (1) of Proposition 2, we have $\mathbf{C}_{B,(t,T],k}\mathbf{M}_3 = \mathbf{P}_{t+k}(\mathbf{A}^\top)^k$. Given $\mathbf{M}_3 = \hat{\mathbf{M}}_3$ and let $k = 1$ we get $\mathbf{P}_{t+1}\mathbf{A}^\top = \mathbf{P}_{t+1}\hat{\mathbf{A}}^\top$, then $\mathbf{A} = \hat{\mathbf{A}}$ holds. Now for all $k > 1$, we have $\mathbf{P}_{t+k}(\mathbf{A}^\top)^k = \hat{\mathbf{P}}_{t+k}(\mathbf{A}^\top)^k$, then $rank(\mathbf{A}) = r$ implies $\mathbf{P}_{t+k} = \hat{\mathbf{P}}_{t+k}$ for all $k > 1$. Similarly, applying (2) of Proposition 2 we have $\mathbf{C}_{F,[1,t),k}\mathbf{M}_1 = \mathbf{P}_{t-k}(\tilde{\mathbf{A}})^k$. Given $\mathbf{M}_1 = \hat{\mathbf{M}}_1$, $\mathbf{A} = \hat{\mathbf{A}}$, $\boldsymbol{\pi} = \hat{\boldsymbol{\pi}}$, and $rank(\mathbf{A}) = r$, then $\mathbf{P}_{t-k} = \hat{\mathbf{P}}_{t-k}$ follows for all $0 < k < t$. Thus, HMM parameters are all identified, this completes the proof.

https://doi.org/10.1017/psy.2025.10024 Published online by Cambridge University Press

B. Proof of Theorem 2

We extend the notation for the Khatri–Rao product for two matrices to the case with an arbitrary collection of matrices.

Definition 14. The Khatri–Rao matrix product of n matrices $\mathbf{U}_i = (\mathbf{u}_{i1}, \ldots, \mathbf{u}_{ir}) \in \mathbb{R}^{q_i \times r}$ for $i \in [n]$ is denoted by

$$\bigotimes_{i=1}^{n} \mathbf{U}_i = \left(\bigotimes_{i=1}^{n} \mathbf{u}_{i1}, \ldots, \bigotimes_{i=1}^{n} \mathbf{u}_{ir} \right). \tag{A18}$$

Remark 12. Note the product in Definition 14 produces a $\left(\prod_{j=1}^{n} q_i \right) \times r$ matrix.

Example 10. Suppose there are m items, $Y_i \in [q_i]$ for $i \in [m]$, such that the responses are conditionally independent given Z. The conditional probability for a particular response pattern $Y_1, \ldots, Y_m$ is $\mathrm{P}(Y_1 = j_1, \ldots, Y_m = j_m \mid Z = z) = \prod_{i=1}^{m} p_{iz,j_i}$. Therefore, the conditional independence assumption implies that the Khatri–Rao product can be used to form the $q_1 q_2 \cdots q_m \times r$ matrix of response probabilities for $Y_1, \ldots, Y_m$ as

$$\mathbf{P}_{12\cdots m} = \bigotimes_{i=1}^{m} \mathbf{P}_i. \tag{A19}$$

Lemma 5 (Allman et al., 2009, Lemma 13). *Let* $\mathbf{U}_i \in \mathbb{R}^{q_i \times r}$, $q = \prod_{i=1}^{n} q_i$, *and* $\mathbf{U} = \bigotimes_{i=1}^{n} \mathbf{U}_i$. *Then for generic* $\mathbf{U}_i$*'s,* $rank_K(\mathbf{U}) = rank(\mathbf{U}) = \min(q, r)$.

Using Lemma 5, we can establish our Theorem for the generic identifiability of heterogeneous HMMs.

We first need to show that $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ from Equation A17 generically satisfy Kruskal's theorem. For generic choice of $\mathbf{A} = \mathbf{I}_r$, these $\mathbf{M}_1$ and $\mathbf{M}_3$ reduce to $\widetilde{\mathbf{M}}_1 = \mathbf{P}_1 * \mathbf{P}_2 * \cdots * \mathbf{P}_{t-1}$, $\mathbf{M}_2 = \mathbf{P}_t \mathbf{D}_\pi$, and $\widetilde{\mathbf{M}}_3 = \mathbf{P}_{t+1} * \cdots * \mathbf{P}_T$. The $\mathbf{P}_t$'s are stochastic matrices so we can express them as $\mathbf{P}_t = \mathbf{U}_t \mathbf{D}_t$ where $\mathbf{D}_t$ is a diagonal matrix of non-zero scalars defined so that the columns sum to one. Recall that both the rank and Kruskal rank are unaffected by multiplying columns by nonzero scalars, so we can apply Lemma 5 to a Khatri–Rao product of $\mathbf{P}_1, \ldots, \mathbf{P}_T$. Therefore, by applying Lemma 5, we have $rank_K(\widetilde{\mathbf{M}}_1) = rank_K(\mathbf{M}_2) = rank_K(\widetilde{\mathbf{M}}_3) = r$, and then Theorem 4 generically holds. Thus by Theorem 4, from the joint distribution of the heterogeneous HMM for generic $\mathbf{A}$, $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{M}_3$ are uniquely determined up to permutations of hidden states.

By employing a similar proof technique as in Theorem 1, and transition matrix $\mathbf{A}$ is generically of rank r, we can then unravel $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ to identify the heterogeneous HMM parameters $\mathbf{P}_1, \ldots, \mathbf{P}_T$, $\mathbf{A}$, and $\boldsymbol{\pi}$.

C. Proof of Corollary 1

The proof follows by recalling Chen et al. (2020) showed that if $\mathbf{\Delta}_t$ and $\mathbf{\Delta}_{t-k}$ are of the form $(\mathbf{D}_1^\top, \mathbf{\Delta}'^\top)^\top$ with $\mathbf{D}_1 \in \mathbb{D}_s$ then $rank_K(\mathbf{P}_t) = rank_K(\mathbf{P}_{t-k}) = 2^K$, so the conditions in Assumption 1 are satisfied and Theorem 1 can be applied, which proves the corollary.

D. Proof of Corollary 2

Chen et al. (2020, Theorem 1) establishes that the condition (1) ensures that the column-rank of $\mathbf{P}_t$ and $\mathbf{P}_{t-k}$ are generically 2^K. Also, $\mathbf{A}$ is generically full rank, so the conditions for Assumption 1 hold generically and Theorem 1 can be applied, which completes the proof.

E. Proof of Theorem 3

The proof follows closely to the proof of Theorem 1 with an exception being the condition that $rank_K(\mathbf{P}_{11}) \geq 2$. First, note that $\mathbf{P}_{12} = \mathbf{P}_2$ and $rank_K(\mathbf{P}_{12}) = r$ imply that $rank_K(\mathbf{B}_{(1,T]}) = r$. Consequently,

$$\mathrm{rank}_K(\mathbf{P}_{11}\mathbf{D}_{\pi_1}) + \mathrm{rank}_K(\mathbf{P}_{12}) + \mathrm{rank}_K\left(\mathbf{B}_{(1,T]}\right) \geq 2(r+1), \tag{A20}$$

so the $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ defined in Equation 10 are unique. The uniqueness of $\mathbf{M}_1$ implies that $\mathbf{P}_{11}\mathbf{D}_{\pi_1}$ is uniquely determined. Pre-multiplying $\mathbf{P}_{11}\mathbf{D}_{\pi_1}$ by $\mathbf{1}_{q_{11}}^\top$ implies that $\boldsymbol{\pi}_1$ is unique, so that $\mathbf{P}_{11}$ is also uniquely determined. The uniqueness of $\mathbf{M}_2$ directly implies $\mathbf{P}_{12}$ is identified. As described in the proof of Theorem 1, the uniqueness of $\mathbf{P}_{12}$ and $\mathbf{M}_3$ together imply that terms making up $\mathbf{B}_{(1,T]}$ can be unraveled so that $\{\mathbf{P}_t\}_{t>1}$ and $\mathbf{A}$ are identified.

F. Proof of Corollary 3

The proof follows by recalling Chen et al. (2020) showed that if $\mathbf{\Delta}_{12} = (\mathbf{D}_1^\top, \mathbf{\Delta}'^\top)^\top$ with $\mathbf{D}_1 \in \mathbb{D}$ and $\mathbf{\Delta}_{11}$ satisfies Assumption (2b) then $rank_K(\mathbf{P}_{12}) = 2^K$ and $rank_K \mathbf{P}_{11} \geq 2$, respectively. The conditions of Theorem 3 are satisfied, which proves the corollary.

G. Proof of Corollary 4

Chen et al. (2020, Theorem 1) establishes that the condition (1) ensures that the column-rank of $\mathbf{P}_{12}$ is generically full rank and $\mathbf{P}_{11}$ has a Kruskal rank of at least two. Also, $\mathbf{A}$ is generically full rank, so the conditions in Theorem 3 generically hold, which completes the proof.

Cite this article: Liu, Y. and Culpepper, S. (2025). Designing Learning Intervention Studies: Identifiability of Heterogeneous Hidden Markov Models. *Psychometrika*, 1258–1283. https://doi.org/10.1017/psy.2025.10024