This is a "preproof" accepted article for *Psychometrika*.
This version may be subject to change during the production process.
DOI: 10.1017/psy.2025.10038

Explaining Performance Gaps with Problem-Solving Process Data via Latent Class Mediation Analysis

Sunbeom Kwon and Susu Zhang

University of Illinois Urbana-Champaign

Abstract

Process data, in particular, log data collected from a computerized test, documents the sequence of actions performed by an examinee in pursuit of solving a problem, affording an opportunity to understand test-taking behavioral patterns that account for demographic group differences in key outcomes of interest, for instance, final score on a cognitive item. Addressing this aim, this paper proposes a latent class mediation analysis procedure. Using continuous process features extracted from action sequence data as indicators, latent classes underlying the test-taking behavior are identified in a latent class mediation model, where an examinee's nominal latent class membership enters as the mediator between the observed grouping and outcome variables. A headlong search algorithm for selecting the subset of process features that maximizes the total indirect effect of the latent class mediator is implemented. The proposed procedure is validated with a series of simulations. An application to a large-scale assessment highlights how the proposed method can be used to explain performance gaps between students with learning disability and their typically developing peers on the National Assessment of Educational Progress (NAEP) math assessment.

Keywords: process data, latent class analysis, mediation analysis, variable selection, large-scale assessment

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Explaining Performance Gaps with Problem-Solving Process Data via Latent Class Mediation Analysis

Introduction

Using computers as assessment delivery platforms allowed the collection of process data, which is computer log data that documents an examinee's sequence of actions (e.g., clicks, keystrokes, and revisits) while solving a task (Bergner & von Davier, 2019). Typically, the sequence of actions of an examinee on a particular item is stored as a tuple of nominal elements, each representing a specific action. For example, an action sequence on a constructed response item might be: (Enter_Item, Open_Scratchwork, Draw, Clear, Zoom_In, Type_7.35, Exit_Item, Enter_Item, Type_73.5, Exit_Item). It shows us what tools the examinee utilized, what answers the examinee typed in before submitting the final response, and how many times the examinee visited this item page on the computer. Such data can preserve valuable information on how examinees arrived at their outcome, thus providing information beyond response data (i.e., correct/incorrect). A rich body of literature demonstrated the utility of process data for common measurement and educational tasks, for instance, to build measurement models characterizing examinee and item characteristics (e.g., Chen et al., 2020; Fang, Ying, et al., 2020; LaMar, 2018; Xiao and Liu, 2024; Zhan and Qiao, 2022) and improve proficiency scoring (e.g., He, Shi, and Tighe, 2023; Zhang et al., 2022), to identify behavioral prototypes or stages of problem-solving (e.g., Eichmann, Greiff, et al., 2020; Hao and Mislevy, 2019; He et al., 2019, 2022; Tang, 2023; Ulitzsch, He, and Pohl, 2022; Wang et al., 2020), and to identify behavioral characteristics that predict final performance (e.g., Greiff et al., 2015; He and von Davier, 2016; Qiao and Jiao, 2018; Ulitzsch, Ulitzsch, et al., 2022; Ulitzsch et al., 2021).

The current paper focuses on using process data to understand problem-solving patterns that account for group differences in test scores. Test scores play a vital role in many key decisions, both for individual candidates (e.g., in college admissions, licensing, and recruitment) and for educators and policymakers using formative and large-scale

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

assessment data to guide instruction and policy development. Understanding demographic subgroup differences in test-taking behavior and performance is critical for mitigating potential test biases and closing achievement gaps. An example is the achievement gap in mathematics between US students from underrepresented groups, such as racial minority groups and students with disabilities, and their peers, which has been persistently reported based on the National Assessment of Educational Progress (NAEP) over the years (U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, 2022). While the NAEP assessments are designed to measure student performance instead of to explain the differences, there is growing interest in the potential utility of test-taking process data, coupled with student background and proficiency information, to provide additional insights into how problem-solving behavior (e.g., test-taking strategies, misconceptions, use of accommodation/universal design tools) explains performance differences across demographic groups. This is exemplified by the release of the restricted-use process data from select blocks of the NAEP 2017 Grade 8 and Grade 4 math assessments (NCES, 2020), as well as recent Institute of Education Sciences (IES) calls for proposals on the use of NAEP process data to understand the link between test-taking behavior and mathematics performance for learners with disabilities, the goal being to gather evidence that ultimately contributes to the improvement of learning of these students from special populations.

Indeed, many previous studies have shown that analyzing process data can aid in understanding subgroup differences (e.g., He and von Davier, 2016; Liao et al., 2019) and explaining differences in sequential patterns in correct/incorrect problem-solving (e.g., Greiff et al., 2015; He and von Davier, 2016; Ulitzsch, He, and Pohl, 2022). While these findings provide supporting evidence on the potential use of process data to understand subgroup differences in item performance, the limitation of prior approaches for investigating the process data is that the relationship between action sequence patterns and demographic backgrounds (e.g., Eichmann, Goldhammer, et al., 2020), and similarly the

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

relationship between action sequence patterns and final response (e.g., Eichmann, Greiff, et al., 2020; Gao, Zhai, et al., 2022; He, Borgonovi, and Suárez-Álvarez, 2023) are studied separately. This does not directly address the question of what types of sequential patterns contributed to group differences. Addressing this requires modeling problem-solving patterns as a potential mediator that explains group differences in the final response. To date, no model-based approach directly addresses this need. We propose a latent class mediation analysis (LCMA) procedure to address this question. Using continuous process features extracted from action sequence data (e.g., features extracted using multidimensional scaling) as indicators, latent classes underlying the test-taking behavior are identified in a latent class mediation model, where an examinee's nominal latent class membership enters as the mediator between the observed grouping and outcome variables.

In the traditional latent variable mediation analysis, the mediator is a continuous latent construct that mediates the predictor's effect on the outcome in a linear fashion. Two methods can be used to estimate the mediation effect: the difference in coefficients method and the product of coefficients method. In the difference in coefficients method, an outcome is regressed on the predictor and then on both the predictor and mediator, and the indirect effect is the difference in the coefficient of the predictor. In the product of coefficients method, the mediator is regressed on the predictor, and the outcome is regressed on the predictor and mediator, and the indirect effect is the product of the coefficients associated with the predictor-mediator and mediator-outcome relationships. By contrast, in latent class mediation analysis, the mediator is a discrete grouping variable whose membership probabilities change with the predictor and generate stepwise changes in the outcome. When both the mediator and outcome are continuous, the total effect of the predictor can be additively decomposed into direct and indirect effects. However, this additive decomposition is not straightforward when the mediator is discrete, and traditional methods for identifying indirect effects are no longer applicable (Sint et al., 2021). A counterfactual framework (Pearl, 2010; Robins & Greenland, 1992) resolves these

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

issues by defining direct effect and total indirect effect for discrete mediators. The total indirect effect summarizes the mediation effect of a latent class mediator as the expected outcome difference in a focal group when class membership changes from what it would be under the focal group to what it would be under a reference group.

One difficulty in analyzing process data arises from the nonstandard format of response processes. That is, the length of action sequences varies across examinees and is coded as nominal elements, making traditional analyses inapplicable to process data such as generalized linear models. Addressing the issue of unstructured data format, we work with features extracted from process data. One example of a process feature extraction method is multidimensional scaling (MDS; Borg and Groenen, 2005; Tang et al., 2020). The extracted MDS features are in a rectangular data format and scaled on a continuum while containing the information of the original action sequences, making it suitable for the proposed LCMA procedure.

Another challenge in process data analysis is that the features extracted from the process data are often high-dimensional. To address this issue, we further perform dimension reduction of the process features via model-based clustering on the process features, that is, latent class analysis. Latent class analysis (Banfield & Raftery, 1993; Lazarsfeld, 1950; Lazarsfeld, 1968; Oberski, 2016; Vermunt & Magidson, 2002) can be used to identify latent nominal variables through a set of observed indicators. Clustering is often used to explore common sequential patterns and to link them to variables of interest, such as final performance and demographics (e.g., Gao, Cui, et al., 2022; Hao and Mislevy, 2019; He, Borgonovi, and Suárez-Álvarez, 2023). Here, we use the term latent class to refer to the latent profile or the Gaussian mixture component underlying continuous indicators. Identifying latent classes in process data can classify examinees into subgroups based on their test-taking behavior and reveal individual differences in sequential patterns (e.g., Bergner and von Davier, 2019; Welling et al., 2024).

These latent classes may also help explain performance gaps, such as those observed

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

on the NAEP Math Assessment between students with learning disabilities and their peers (Judge & Watson, 2011). This can be achieved by considering the latent class variable as a mediator explaining the effect of a predictor on the outcome (e.g., Muthén, 2011; Sint et al., 2021). Literature discussing latent classes as potential mediators has primarily focused on latent class mediators with discrete indicators (e.g., Hsiao et al., 2021; Muthén, 2011). However, there is a lack of methodological investigation in latent class mediation analysis with continuous indicators (Hsiao et al., 2021). Literature considering the extension of latent class analysis with continuous indicators is limited to latent class model with either covariates (Murphy & Murphy, 2020; Vermunt & Magidson, 2002), or latent class model with distal outcomes (Dziak et al., 2016; Vermunt, 2010). In this study, we extend the latent class analysis with continuous indicators (e.g., process features) to explain the effect of a binary predictor on a binary outcome through the nominal latent class mediator. An EM algorithm is implemented for parameter estimation.

Extracted process features may contain noise or irrelevant information, which can weaken the generalizability of results in latent class mediation models. Removing noisy indicators can enhance classification accuracy and parameter precision in latent class analysis (Dean & Raftery, 2010). To address this, variable selection methods, such as the headlong search algorithm, have been proposed to identify the optimal set of indicators. In this study, a headlong search algorithm, which is generally used to explore the model space and select clustering variables, was used to select process features that maximize the total indirect effect of the latent class mediator in explaining group differences in outcomes.

In summary, we propose a latent class mediation analysis procedure for 1) identifying the latent class underlying the distribution of process features, 2) finding the set of process features that can best explain the effect of observed group membership on the outcome, and 3) assessing the indirect effect of the group membership on the outcome through the nominal latent class mediator. A headlong search algorithm is used to find the set of process features that best explains the group difference in performance. This is

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

achieved by finding the optimal subset of process features that maximizes the total indirect effect. The proposed framework is intended primarily as an exploratory tool for hypothesis generation from the complex process data, rather than a confirmatory tool for drawing causal conclusions about test-taking behaviors.

The rest of the article is structured as follows. The next section begins with a motivating example based on one item from the NAEP 2017 grade 8 math assessment. Then, the latent class mediation model and the parameter estimation algorithms are introduced. It is followed by the headlong search algorithm for selecting the optimal set of process features. In a simulation study, the performance of the proposed analysis procedure is evaluated in terms of classification accuracy and parameter estimation accuracy. This is followed by an empirical application of the procedure on the NAEP Math Assessment item from the motivating example. Lastly, the significance and limitations of the current study are discussed.

Latent Class Mediation Analysis

Motivating Example

As a motivating example, we consider one item available from the restricted-use response and process data in the digital version of the 2017 NAEP Grade 8 Math Assessment. NAEP adopts a probabilistic sampling approach to select schools and students to represent the diverse student population in the United States. The data set consisted of $28,194$ nationally sampled students who were administered a 15-item block (block 1717MA2N03CLID30EX) on the eNAEP, which was administered with a Surface tablet and a stylus. The eNAEP was also embedded with a set of universal design tools, including scratchwork (where students could draw and erase), zooming, color theme change, equation editor, text-to-speech (TOS), and highlighting. Students were allowed to revisit an item multiple times during the test, and each enter/exit of the item page was recorded. For this block, students were not allowed to use a calculator. The data set consisted of students' ordinal scores to the 15 math items, as well as their log data on the math block, which

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

contained student interactions with the eNAEP platform, such as item visits, tool usage, and response entries to the 15 multiple choice, constructed response, or drag-and-drop items. Students, teachers, and schools also completed a series of survey questionnaires, which contained information on students' disability status and accommodation on the test.

In the NAEP Math Assessment, students with learning disabilities (LD) consistently underperformed compared to their typically developing (TD) peers (Judge & Watson, 2011). For the current example, we aim to identify test-taking process patterns that can explain this performance gap between LD and TD learners, by focusing on one item on the multiplication of decimals (VH336968) from this block (Figure 1). The item asked students to find the solution to 1.5×4.9 without using a calculator, and the correct response was 7.35. This item was chosen because it was a constructed response item allowing various responses, and it was a relatively computationally involving task, where students use a certain tool (i.e., scratchwork) to facilitate computation.

Figure 1

Item VH336968 from the 2017 National Assessment of Educational Progress (NAEP) Grade 8 Math Assessment.

Note. https://www.nationsreportcard.gov/nqt/

The NAEP restricted-use log data recorded each response entry to a constructed response item, from putting the cursor in the textbox to leaving the textbox, as one event. The log data thus contained the sequence of interactions of a student on the item, including various constructed response entries (a student can have multiple entries if they

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

made answer changes throughout the test), tool usage, and item revisits (Exit_Item, Enter_Item in the middle of the action sequence). In the data preprocessing stage, we removed system events from the log data and recoded repeated actions, such as consecutive draws/erases for each stroke using the scratchwork tool, into a single action. The first and the last actions (Enter_Item, Exit_Item) were discarded as these were the common elements in all students' action sequences. We masked the final responses to ensure the action sequence does not directly predict the final outcome, and the answer entries were recoded into two categories. The "735" category includes answers containing the number sequence 7, 3, and 5, with the decimal place masked. The "non-735" category includes responses that do not include the numbers 7, 3, or 5. A preliminary analysis revealed a common error where many test takers placed the decimal point incorrectly, leading to errors of 735 and 73.5. Recoding the answers in this way masked the final responses while retaining information about the types of mathematical concepts the test takers struggled to demonstrate. Students with disabilities other than LD (e.g., Autism) and those who received extended time accommodation (90-minute version) were excluded from the analysis. The sample size of the LD group was 590. 2500 students from the TD group were randomly selected to balance the sample size between the two groups and reduce computational demand. The sample size of the final data set used in the analysis was thus $N = 3090$. Descriptive statistics of the sample are given in Table 1. The marginal proportion of correct responses was .49 for the TD group and .21 for the LD group.

To transform the process data into continuous features suitable for the subsequent analysis while preserving the original sequential pattern information, multidimensional scaling (MDS) was applied for feature extraction. MDS is a dimension reduction method that extracts latent features based on the pairwise dissimilarity measure between two observations. The technical details of extracting MDS features from the action sequence process data are summarized in Appendix A. The proposed LCMA procedure has a multivariate normal distributional assumption on the indicators of the latent class variable.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Table 1

Descriptive Statistics of the NAEP Math Assessment Item VH336968

	LD	TD		LD	TD
Response time (secs)	102.55	88.03	Disability severity		
Male, %	64	49	Profound, %	3	
Age	14.57	14.38	Moderate, %	29	
White, %	47	47	Mild, %	59	
African American, %	11	14	Omitted, %	8	
Hispanic, %	25	24	Breaks during test, %	7	0
Other, %	16	14	Cueing, %	3	0
ELL, %	13	4	Bilingual dictionary, %	1	0
			Preferential seating, %	5	0
			Separate sessions, %	13	0

Note. ELL is the English language learners. The number of LD students was 590 and the number of TD students was 2500. The sample sizes are rounded to the closest 10.
SOURCE: U.S. Department of Education, National Center for Education Statistics, "Response Process Data from the NAEP 2017 Grade 8 Mathematics Assessment."

The process features extracted from MDS are scaled on a continuum and are suitable for the proposed analysis. However, note that our proposed method is not limited to process features from MDS. Any feature extraction method that transforms the original action sequence data to a rectangular and continuous data format while preserving the information of examinees' problem-solving behavior could serve as a viable alternative to MDS. Based on a five-fold cross-validation, $K = 15$ total features were extracted. The cross-validation was run on the dissimilarity matrix of the action sequence data using the `ProcData` R package (Tang et al., 2021). The dissimilarity matrix of the action sequence data was obtained as described in Appendix A. Then, the 15 process features M_k $(k = 1, ..., K)$ extracted using MDS were used as the potential candidates of continuous indicators in the latent class mediation analysis.

The latent class mediation analysis aims to find the latent classes underlying the process features that can explain the correct response probability gap between the LD and TD students. In the latent class mediation model, the predictor G was the binary disability

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

membership variable, where $G = 0$ if the student belongs to the typically developing group and $G = 1$ if the student belongs to the learning disability group, the outcome Y was the binary score on the multiplication item, with $Y = 0$ indicating an incorrect response, and $Y = 1$ indicating a correct response (i.e., answers equivalent to 7.35). The English language learner (ELL) variable was included as a covariate X to control for potential confounding effects between the predictor and mediator, as well as between the mediator and outcome. Here, $X = 1$ indicates an English language learner, and $X = 0$ indicates otherwise. The $K = 15$ process features, $\mathbf{M}$, were the candidate indicators of the latent class membership variable (Ω) that mediates the relationship between G and Y. The proposed latent class mediation analysis procedure can be applied to find the optimal subset of process features maximizing the total indirect effect of the latent class mediator between the predictor and the outcome. We next articulate the model formulation as well as the technical details.

Latent Class Mediation Model

The latent class part of the model assumes a nominal latent class variable Ω_i $(i = 1, ..., N)$ for N observations exists underlying the distribution of relevant process features $\mathbf{M}_{\kappa,i}$. The set of relevant features $\mathbf{M}_{\kappa,i}$ is assumed to follow a mixture of multivariate normal distributions with class-specific mean $\boldsymbol{\mu}_\omega$ and covariance $\boldsymbol{\Sigma}_\omega$.

$$\mathbf{M}_{\kappa,i} \mid \Omega_i = \omega \sim \text{MVN}(\boldsymbol{\mu}_\omega, \boldsymbol{\Sigma}_\omega). \tag{1}$$

Equation (1) implies that the distribution of an examinee's process features, which contain information on their sequential patterns in pursuit of solving the item, differs across the latent classes. For a randomly sampled examinee, the probability density function of $\mathbf{M}_{\kappa,i}$ given $\underset{\sim}{\boldsymbol{\mu}} = \{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_L\}, \underset{\sim}{\boldsymbol{\Sigma}} = \{\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_L\}$, and $\boldsymbol{\pi} = \{\pi_1, ..., \pi_L\}$ is

$$f(\mathbf{M}_{\kappa,i}|\underset{\sim}{\boldsymbol{\mu}}, \underset{\sim}{\boldsymbol{\Sigma}}, \boldsymbol{\pi}) = \sum_{l=1}^{L} \pi_l f_l(\mathbf{M}_{\kappa,i} \mid \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l), \tag{2}$$

where L is the number of latent classes and π_l is the probability of belonging to latent class l. Here, $f_l(\mathbf{M}_{\kappa,i}|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ denotes the class-specific multivariate normal density.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

The effect of the binary group membership variable G_i on the latent class Ω_i, controlling for covariate X_i $(i = 1, ..., N)$, can be described by a multinomial logistic regression model in Equation 3.

$$P(\Omega_i = \omega \mid G_i = g, X_i = x) = \frac{e^{\beta_{0\omega}+\beta_{1\omega}g+\xi_\omega x}}{\sum_{l=1}^{L} e^{\beta_{0l}+\beta_{1l}g+\xi_l x}}. \tag{3}$$

The regression coefficients $\beta_{0\omega}$, $\beta_{1\omega}$ and ξ_ω are the class-specific intercept and slope for class ω. For model identification, we set the intercept and slope of the first class to $\beta_{01} = \beta_{11} = \xi_1 = 0$. Equation (3) implies that for an examinee i, the membership probability associated with the problem-solving latent class, Ω_i, depends on the observed group membership G_i, controlling for the covariate X_i. When the predictor G does not represent a randomized intervention, associations among variables may be influenced by confounding factors. In such cases, it is common practice to adjust for potential confounders of the predictor–mediator $(G \to \Omega)$ and mediator–outcome $(\Omega \to Y)$ associations by including relevant covariates in the model (Muthén, 2011; Preacher, 2015; Valente et al., 2017; Witkiewitz et al., 2018). This approach helps to reduce bias in the estimated associations.

Given the group membership G_i and the latent class membership Ω_i, examinee i's outcome Y_i is modeled via a logistic model, controlling for the covariate X_i,

$$P(Y_i = 1 \mid G_i = g, \Omega_i = \omega, X_i = x) = \frac{e^{\gamma g+\alpha_\omega+\zeta x}}{1 + e^{\gamma g+\alpha_\omega+\zeta x}}. \tag{4}$$

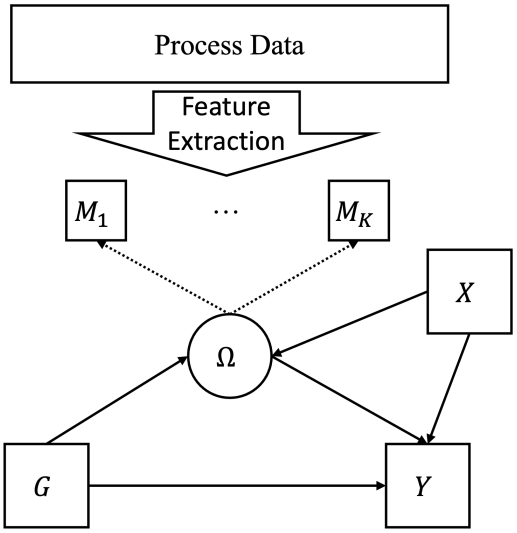
Each latent class of the problem-solving process is associated with a class-specific intercept (α_ω). The coefficient vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_L)'$, together with $\boldsymbol{\beta}_0 = (\beta_{01}, ..., \beta_{0L})'$ and $\boldsymbol{\beta}_1 = (\beta_{11}, ..., \beta_{1L})'$, are associated with the indirect effect of the group membership G on the outcome Y, mediated by the nominal latent class Ω. The coefficient γ is associated with the direct effect of G on Y, after controlling for Ω and covariate X. Figure 2 shows the structure of the latent class mediation model using process data.

The likelihood of the model parameters given the observed group memberships $\mathbf{G} = (G_1, \ldots, G_N)'$, the final outcome $\mathbf{Y} = (Y_1, ..., Y_N)'$, the process features

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Figure 2

Latent Class Mediation Model



Note. M_k *represents a process feature,* Ω *is a latent class variable,* G *is a binary group membership (e.g., learning disability = 1 vs. typically developing = 0),* Y *is a binary outcome (e.g., correct = 1 vs. incorrect = 0)*

, and X is a covariate. Solid arrows indicate predictive relationships: G and X predict Ω, while Ω and X predict Y. The dashed arrows indicate that the M_ks serve as measurement indicators of Ω.

$\mathbf{M}_\kappa = (\mathbf{M}_{\kappa,1}, ..., \mathbf{M}_{\kappa,N})'$, and the covariate $\mathbf{X} = (X_1, ..., X_N)'$ is

$$
\begin{aligned}
&L(\underset{\sim}{\boldsymbol{\mu}}, \underset{\sim}{\underset{\sim}{\boldsymbol{\Sigma}}}, \gamma, \boldsymbol{\alpha}, \zeta, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\xi}; \mathbf{Y}, \mathbf{M}_\kappa, \mathbf{X}) \\
&\quad = P(\mathbf{Y}, \mathbf{M}_\kappa, \mathbf{X} \mid \mathbf{G}, \underset{\sim}{\boldsymbol{\mu}}, \underset{\sim}{\underset{\sim}{\boldsymbol{\Sigma}}}, \gamma, \boldsymbol{\alpha}, \zeta, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\xi}) \\
&\quad = \prod_{i=1}^{N} \sum_{l=1}^{L} P(\Omega_i = l \mid g_i, x_i, \beta_{0l}, \beta_{1l}, \xi_l) P(Y_i = y_i \mid g_i, x_i, \gamma, \alpha_l, \zeta) P(\mathbf{M}_{\kappa,i} \mid \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \\
&\quad = \prod_{i=1}^{N} \sum_{l=1}^{L} \left\{ \frac{e^{\beta_{0l} + \beta_{1l} g_i + \xi_l x_i}}{\sum_{d=1}^{L} e^{\beta_{0d} + \beta_{1d} g_i + \xi_d x_i}} \times \frac{e^{y_i(\gamma g_i + \alpha_l + \zeta x_i)}}{1 + e^{\gamma g_i + \alpha_l + \zeta x_i}} \times f_l(\mathbf{M}_{\kappa,i} \mid \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \right\}.
\end{aligned}
\tag{5}
$$

Note that the process features are assumed to be independent of the final outcome given the latent class membership. That is, the latent class is assumed to fully capture the relationship between the process features and the outcome, given the covariates.

The number of latent classes (L) is determined by fitting the latent class model only

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

using the process features and comparing the Bayesian information criterion (BIC).

$$BIC = 2 \times loglikelihood - p \times log(N), \tag{6}$$

where p is the number of parameters, and N is the sample size. BIC is known to be consistent in choosing the number of classes in a mixture model (Keribin, 1998).

The class-specific covariance matrix of process features for class l, $\mathbf{\Sigma}_l$, is parameterized through an eigenvalue decomposition of the following form:

$$\mathbf{\Sigma}_l = \lambda_l \mathbf{D}_l \mathbf{A}_l \mathbf{D}_l^T, \tag{7}$$

where λ_l is a scalar controlling the volume of the ellipsoid, $\mathbf{A}_l$ is a diagonal matrix specifying the shape with $|\mathbf{A}_l| = 1$, and $\mathbf{D}_l$ is an orthogonal matrix determining the orientation of the ellipsoid (Banfield & Raftery, 1993; Celeux & Govaert, 1995; Fraley & Raftery, 2002). Various equality constraints can be assumed between and within group covariance structures. In their works, Banfield and Raftery (1993) and Celeux and Govaert (1995) present models tailored to various clustering scenarios. These models are implemented in the `mclust` R package (Scrucca et al., 2023). Celeux and Govaert (1995) recommended using the model allowing different volumes and more parsimonious models, such as a diagonal covariance matrix for high-dimensional data. Here, we adopted the model that assumes varying volumes but equal shapes between classes and orientations aligned with the coordinate axes. In this parsimonious model, the class-specific covariance matrix becomes,

$$\mathbf{\Sigma}_l = \lambda_l \mathbf{B}, \tag{8}$$

where $\mathbf{B}$ is a diagonal matrix with $|\mathbf{B}| = 1$.

Parameter Estimation

An EM algorithm (Dempster et al., 1977) is implemented to find the marginal maximum likelihood estimates of the latent class mediation model by maximizing the observed data log-likelihood. Similar to the EM algorithm for the Gaussian mixture model

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

in Fraley and Raftery (2002), the class membership variable $\mathbf{Z}_i = (Z_{i1}, ..., Z_{iL})$ is introduced as the unobserved portion of the data, where,

$$Z_{il} = \begin{cases} 1, & \text{if } \Omega_i = l, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The conditional distribution of $(Y_i, \mathbf{M}_{\kappa,i}, X_i)$ given $\mathbf{Z}_i$ is

$$\prod_{l=1}^{L} [P(\Omega_i = l \mid g_i, x_i, \beta_{0l}, \beta_{1l}, \xi_l) P(Y_i \mid g_i, x_i, \gamma, \alpha_l, \zeta) P(\mathbf{M}_{\kappa,i} \mid \mu_l, \mathbf{\Sigma}_l)]^{Z_{il}}. \tag{10}$$

The log-likelihood of the parameters given the complete data $\underset{\sim}{\mathbf{U}} = (Y_i, \mathbf{M}_{\kappa,i}, X_i, \mathbf{Z}_i)_{1 \le i \le N}$ is,

$$\begin{aligned} &\ell_c(\underset{\sim}{\boldsymbol{\mu}}, \underset{\sim}{\mathbf{\Sigma}}, \gamma, \boldsymbol{\alpha}, \zeta, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\xi}; \underset{\sim}{\mathbf{U}}) \\ &= \sum_{i=1}^{N} \sum_{l=1}^{L} \Big\{ Z_{il} \Big[\log\, P(\Omega_i \mid g_i, x_i, \beta_{0l}, \beta_{1l}, \xi_l) \\ &\quad + \log\, P(Y_i \mid g_i, x_i, \gamma, \alpha_l, \zeta) + \log\, P(\mathbf{M}_{\kappa,i} \mid \mu_l, \mathbf{\Sigma}_l) \Big] \Big\}. \end{aligned} \tag{11}$$

The initial class memberships in the EM algorithm are obtained by fitting the hierarchical agglomeration clustering analysis (Murtagh & Legendre, 2014). The algorithm iterates the E-step and the M-step described below until a convergence criterion has been reached.

E-step

In the E-step, class membership probabilities, $\hat{Z}_{il}$s, are estimated for $i = 1, ..., N$ and $l = 1, ..., L$ in the rth iteration by

$$\hat{Z}_{il,r} = \frac{P(\Omega_i = l \mid g_i, x_i, \beta_{0l,r-1}, \beta_{1l,r-1}, \xi_{l,r-1}) P(Y_i, \mathbf{M}_{\kappa,i} \mid g_i, x_i, \mu_{l,r-1}, \mathbf{\Sigma}_{l,r-1}, \gamma_{r-1}, \alpha_{l,r-1}, \zeta_{l-1})}{\sum_{d=1}^{L} P(\Omega_i = d \mid g_i, x_i, \beta_{0d,r-1}, \beta_{1d,r-1}, \xi_{d,r-1}) P(Y_i, \mathbf{M}_{\kappa,i} \mid g_i, x_i, \mu_{d,r-1}, \mathbf{\Sigma}_{d,r-1}, \gamma_{r-1}, \alpha_{d,r-1}, \zeta_{r-1})}. \tag{12}$$

M-step

In the M-step, we update the parameters, $\underset{\sim}{\boldsymbol{\mu}}, \underset{\sim}{\mathbf{\Sigma}}, \boldsymbol{\lambda}, \gamma, \boldsymbol{\alpha}, \zeta, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1$, and $\boldsymbol{\xi}$ by maximizing the expected complete data log-likelihood computed with the estimates $\hat{\mathbf{Z}}_{1,r}, ..., \hat{\mathbf{Z}}_{n,r}$. For updating $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$, we set the first latent class as the baseline reference level for identifiability, and

$$\log \frac{P(\Omega_i = l)}{P(\Omega_i = 1)} = \beta_{0l} + g_i \beta_{1l} + x_i \xi_l \quad \forall\, l \in \{2, ..., L\}, \tag{13}$$

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

where $\beta_{11} = \xi_1 = 0$. The $\boldsymbol{\beta}$ and the $\boldsymbol{\xi}$ are updated with the estimates from the multinomial logistic regression model by maximizing

$$\sum_{i=1}^{N}\sum_{l=1}^{L}\hat{Z}_{il}[\log\ P(\Omega_i = l \mid g_i, x_i, \beta_{0l}, \beta_{1l}, \xi_l)]. \tag{14}$$

Similarly, $\boldsymbol{\alpha}$, γ, and ζ are updated by maximizing the following term

$$\sum_{i=1}^{N}\sum_{l=1}^{L}\hat{Z}_{il}[\log\ P(Y_i, \mathbf{M}_{\kappa,i} \mid g_i, \mu_l, \boldsymbol{\Sigma}_l, \gamma, \alpha_l)] \tag{15}$$

The closed-form solutions to Equations (14) and (15) are unavailable, so a quasi-Newton method (BFGS; Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) was used to update $\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \gamma$, and ζ.

The class-specific means on the process features, $\boldsymbol{\mu}_l$s, have closed-form expressions from the E-step as

$$\boldsymbol{\mu}_l = \frac{\sum_{i=1}^{n}\hat{Z}_{il}\mathbf{M}_{\kappa,i}}{n_l}, \tag{16}$$

where $n_l = \sum_{i=1}^{n}\hat{Z}_{il}$. For updating the covariance matrix $\boldsymbol{\Sigma}_l = \lambda_l\mathbf{B}$, we use the approach described in Celeux and Govaert (1995). The scattering matrix $\mathbf{W}_l$ of a class is

$$\mathbf{W}_l = \sum_{i=1}^{n}\hat{Z}_{il}(\mathbf{M}_{\kappa,i} - \overline{\mathbf{M}}_\kappa)(\mathbf{M}_{\kappa,i} - \overline{\mathbf{M}}_\kappa)^T. \tag{17}$$

We update λ_l and $\mathbf{B}$ by minimizing

$$\sum_{l=1}^{L}\frac{1}{\lambda_l}tr(\mathbf{W}_l\mathbf{B}^{-1}) + d\sum_{l=1}^{L}n_l\ln(\lambda_l). \tag{18}$$

The minimization of (18) requires an iterative procedure.

$$\begin{aligned}\lambda_l &= \frac{tr(\mathbf{W}_l\mathbf{B}^{-1})}{dn_l},\\ \mathbf{B} &= \frac{\mathrm{diag}(\sum_{l=1}^{L}\frac{1}{\lambda_l}\mathbf{W}_l)}{|\sum_{l=1}^{L}\frac{1}{\lambda_l}\mathbf{W}_l|^{\frac{1}{d}}},\end{aligned} \tag{19}$$

where d is the dimension of the relevant process features $\mathbf{M}_\kappa$. The E-step and the M-step are iterated until a termination criterion has been reached. Parameter estimates from the last iteration are used as the final estimates. For each examinee, the latent class memberships can be estimated via the maximum a posteriori probability (MAP).

$$\hat{\Omega}_i = \arg\max_l \hat{Z}_{il} \tag{20}$$

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Assessing Direct and Indirect Effect

To quantify the amount of information in the outcome explained by the group membership through the latent class mediator, we adopt the assessment of direct and indirect effects with a nominal mediator described in Muthén (2011). Although intended as an exploratory tool, this model assumes no unmeasured confounding among the predictor, mediator, and outcome, as is standard in causal inference frameworks. Let $Y(g, \omega)$ denote the potential outcome that would have been observed if the group membership was g and the latent class membership was ω for an examinee. The conditional expectation of the outcome Y in group g, when the latent class mediator Ω is held constant at the value it would obtain for group g', controlling for the covariate X, is

$$E[Y(g, \Omega(g'))] = \sum_{l=1}^{L} \{E(Y \mid G = g, \Omega = l) \times P(\Omega = l \mid G = g') \mid X = 0\}. \tag{21}$$

The direct effect (DE) and the total indirect effect (TIE) are defined as follows.

$$\begin{aligned} DE &= E[Y(1, \Omega(0)) - Y(0, \Omega(0)) \mid X = 0], \\ TIE &= E[Y(1, \Omega(1)) - Y(1, \Omega(0)) \mid X = 0]. \end{aligned} \tag{22}$$

The total indirect effect is interpreted as the expectation of the difference between the outcome in the focal group ($G = 1$) when the mediator changes from the values it would obtain in the focal group to the values it would obtain in the reference group ($G = 0$). For example, in the context of the NAEP Math Assessment data, the TIE can be interpreted as the expected difference in the probability of a correct response for LD students when their latent class membership shifts from the class it would take in the LD group to the class it would take in the TD group. The TIE and DE can be estimated with the latent class mediation model parameter estimates. The sum of the direct effect and the total

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

indirect effect is equal to the total effect, $TE = E[Y(1, \Omega(1)) - Y(0, \Omega(0)) \mid X = 0]$.

$$
\begin{aligned}
DE &= \sum_{l=1}^{L}[P(Y=1 \mid G=1, \Omega=l)P(\Omega=l \mid G=0) - P(Y=1 \mid G=0, \Omega=l)P(\Omega=l \mid G=0)] \\
&= \sum_{l=1}^{L}\Big\{\Big(\frac{e^{\alpha_l+\gamma}}{1+e^{\alpha_l+\gamma}} - \frac{e^{\alpha_l}}{1+e^{\alpha_l}}\Big)\frac{e^{\beta_{0\omega}}}{\sum_{d=1}^{L} e^{\beta_{0d}}}\Big\}, \\
TIE &= \sum_{l=1}^{L}[P(Y=1 \mid G=1, \Omega=l)P(\Omega=l \mid G=1) - P(Y=1 \mid G=1, \Omega=l)P(\Omega=l \mid G=0)] \\
&= \sum_{l=1}^{L}\Big\{\frac{e^{\alpha_l+\gamma}}{1+e^{\alpha_l+\gamma}}\Big(\frac{e^{\beta_{0\omega}+\beta_{1\omega}}}{\sum_{d=1}^{L} e^{\beta_{0d}+\beta_{1d}}} - \frac{e^{\beta_{0\omega}}}{\sum_{d=1}^{L} e^{\beta_{0d}}}\Big)\Big\}.
\end{aligned}
\tag{23}
$$

Testing of the total indirect effect is available by constructing confidence intervals using the delta method (Sint et al., 2021) or bootstrap resampling (Muthén, 2011). The approximation of the standard error of TIE using the delta method is described in Appendix B.

Headlong Search Algorithm for Feature Selection

The process features are high-dimensional and may contain noisy information irrelevant to the relationship between the observed group and the final outcome. We implement a headlong search algorithm to find the optimal subset of process features that maximizes the total indirect effect of the latent class variable. Let $\mathbf{M}$ be the set of all K process features. The algorithm starts with an initial subset of process features and iteratively updates the subset (denoted $\kappa \subseteq \{1, \ldots, K\}$), to find an optimal subset κ^* such that the LCMA model with process features $\mathbf{M}_{\kappa^*}$ maximizes the total indirect effect of Ω.

Feature Subset Initialization

We first fit the latent class model using all K process features as indicators from Equation (2). The number of latent classes L_{full} is selected using the Bayesian information criterion (BIC) in Equation (6). Then, we fit a latent class model with a single indicator for each feature in $\mathbf{M}$ with the fixed L_{full}. The average variance of class probability estimates across individuals is calculated, where the class probability estimates $\hat{Z}_{il}$ are calculated in the E-step of the EM algorithm from the single indicator LCA model estimates. The larger

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

the average class probability variance, the indicator gives a better separation of the classes. Similar to the approach in Dean and Raftery (2010), we select $L_{full} - 1$ features with the largest variance of class probabilities as the initial set. Here, $L - 1$ is the maximum number of features needed to identify L latent classes by their locations. With the initial set of features, the latent class mediation model is fit using the current subset, $\mathbf{M}_\kappa$, the group membership variable, G, and the outcome, Y. After selecting the initial set of features, we proceed with the inclusion and exclusion steps of the headlong search algorithm.

Inclusion Step

At any iteration, let $\mathbf{M}_\kappa$ be the subcolumns of $\mathbf{M}$ currently included in the model, and let $\mathbf{M}_{-\kappa}$ be the remaining columns of $\mathbf{M}$ not included in the model. The logic of the inclusion and exclusion steps is that if including a feature in $\mathbf{M}_{-\kappa}$ or excluding a feature from $\mathbf{M}_\kappa$ increases the TIE significantly, then we can add or exclude that feature. In the inclusion step, each process feature in $\mathbf{M}_{-\kappa}$ is a candidate feature. For each candidate feature, the latent class mediation model is fit after adding the feature to $\mathbf{M}_\kappa$. The number of latent classes is determined by selecting the LCA model with the highest BIC. We test if the absolute value of TIE increases significantly after adding the candidate feature by examining whether the 95% confidence interval includes the TIE estimate from the previous subset. The feature that increases the TIE most is added to the current set, $\mathbf{M}_\kappa$, if the increase in TIE is significant. If none of the features increase the TIE significantly when added to the current set, we do not add any feature to $\mathbf{M}_\kappa$.

Exclusion Step

In the exclusion step, the features in $\mathbf{M}_\kappa$ are examined. For each feature in $\mathbf{M}_\kappa$, the latent class mediation model is fit after removing that feature from $\mathbf{M}_\kappa$. The number of latent classes is determined by selecting the LCA model with the highest BIC. The feature that leads to the largest increase in TIE when removed is excluded from $\mathbf{M}_\kappa$ if the 95% confidence interval of the TIE does not contain the TIE estimate from the previous step. If none of the features contribute to a significant increase in TIE when removed from the

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

current set, we do not remove any feature from $\mathbf{M}_\kappa$. If there is no change after a round of inclusion and exclusion steps, the feature set is finalized as $\mathbf{M}_{\kappa^*}$, and the finalized latent class mediation model is fit.

The proposed latent class mediation analysis procedure using process data is summarized in Algorithm 1.

Simulation Study

Simulation studies are conducted to examine whether the proposed procedure selects the signal indicators that effectively explain the mediation effect and accurately estimates the indirect effect of the latent class mediator.

Data Generation

Random samples with $N = 500$ sample size, $L = 4$ latent classes, $K = 10$ indicators, and binary final outcome Y were generated under a latent class mediation model given the binary group membership G generated from a Bernoulli distribution with $p = .5$. The numbers of signal indicators were $S = 5, 3, 1$. The noisy indicators were randomly generated independently of the true latent class membership. Thus, the noisy indicators do not contribute to the classification of subjects into latent classes, and they are irrelevant to the relationship between the predictor G and the outcome Y. Figure 3 presents the true mean structure of the 10 indicators conditioned on the four latent classes, where each column represents a latent class. The first S rows are the mean vectors of the signal indicators. In Figure 4, the distributions of latent classes from one of the simulated data sets are presented on a two-dimensional space using the first two indicators to summarise the simulation conditions. In the $S = 5$ condition, at least three of the signal variables need to be selected to identify the four latent classes by location. In the $S = 3$ condition, all the signal variables must be selected to identify the four latent classes correctly. In the $S = 1$ condition, the first variable ($M1$) is the only indicator we need to identify the four true latent classes. Two levels of class-specific variances were considered, $VAR = 1$ and $VAR = 3$, to control the level of overlap, that is, how much the latent classes can intersect.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Algorithm 1 Headlong Search Algorithm for Feature Selection

Input: $\mathbf{M}_k, \mathbf{Y}, \mathbf{G}, \mathbf{X}$

— **Feature Subset Initialization** —

Fit the full LCA model using all K features to select the number of latent classes L_{full} using BIC

Fit single-indicator LCA model with each K feature with fixed L_{full}

Set the initial feature subset κ by selecting $L_{full} - 1$ features with the largest average class probability variance from single-indicator LCA models

— **Fit Initial LCMA Model** —

Fit LCA model using $\mathbf{M}_\kappa$ to select the number of latent classes $L_{initial}$ using BIC

Fit LCMA model using $\mathbf{M}_\kappa$, $\mathbf{Y}$, $\mathbf{G}$, and $\mathbf{X}$ with $L_{initial}$ and calculate $\widehat{\text{TIE}}_{initial}$

$\widehat{\text{TIE}} \leftarrow \widehat{\text{TIE}}_{initial}$

$(l, u) \leftarrow$ lower and upper bound of 95% C.I of TIE

— **Inclusion and Exclusion Steps** —

while κ remains the same after inclusion and exclusion steps **do**

  — **Inclusion Step** —

  for $k \in \kappa^c$ **do**

    $\kappa^* \leftarrow \kappa \cup \{k\}$

    Fit LCA model using $\mathbf{M}_{\kappa^*}$ to select the number of latent classes L_{κ^*} using BIC

    Fit LCMA model using $\mathbf{M}_{\kappa^*}$, $\mathbf{Y}$, $\mathbf{G}$, and $\mathbf{X}$ with L_{κ^*} and calculate $\widehat{\text{TIE}}_k$

  end for

  $m \leftarrow \arg\max_k |\widehat{\text{TIE}}_k|$

  if $|\widehat{\text{TIE}}_m| > |\widehat{\text{TIE}}|$ and $\widehat{\text{TIE}}_m \notin (l, u)$ **then**

    $\kappa \leftarrow \kappa \cup \{m\}$ {Inclusion}

    $\widehat{\text{TIE}} \leftarrow \widehat{\text{TIE}}_m$

    $(l, u) \leftarrow$ lower and upper bound of 95% C.I of TIE

  end if

  — **Exclusion Step** —

  for $k \in \kappa$ **do**

    $\kappa^* \leftarrow \kappa \setminus \{k\}$

    Fit LCA model using $\mathbf{M}_{\kappa^*}$ to select the number of latent classes L_{κ^*} using BIC

    Fit LCMA model using $\mathbf{M}_{\kappa^*}$, $\mathbf{Y}$, $\mathbf{G}$, and $\mathbf{X}$ with L_{κ^*} and calculate $\widehat{\text{TIE}}_k$

  end for

  $m \leftarrow \arg\max_k |\widehat{\text{TIE}}_k|$

  if $|\widehat{\text{TIE}}_m| > |\widehat{\text{TIE}}|$ and $\widehat{\text{TIE}}_m \notin (l, u)$ **then**

    $\kappa \leftarrow \kappa \setminus \{m\}$ {Exclusion}

    $\widehat{\text{TIE}} \leftarrow \widehat{\text{TIE}}_m$

    $(l, u) \leftarrow$ lower and upper bound of 95% C.I of TIE

  end if

end while

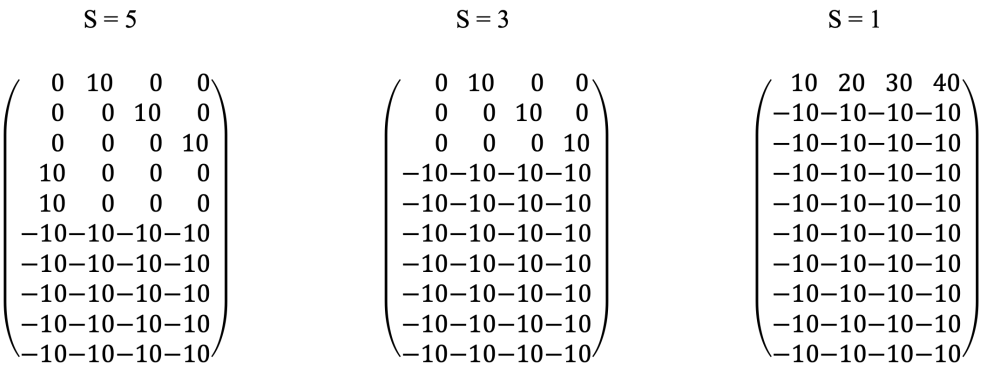
return κ

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Overlapping true latent classes can lead to the misclassification of individuals. In the $VAR = 1$ condition, the latent classes do not overlap, whereas in the $VAR = 3$ condition, the latent classes do overlap, allowing the misclassification of individuals. The true TIE and DE were set to -0.125 and 0. The true model parameter values are described in Appendix C. The number of replications in each condition was $R = 100$. The R codes used for the simulation can be found on the Open Science Framework (OSF) at https://osf.io/a5zem/?view_only=c5b920fee99e4b798006b03880d224ed.

Figure 3

True Mean Structures in the Simulation Study



Note. The columns represent the four latent classes, and the rows represent the ten indicators. The first S rows are the signal indicators, and the rest are the noisy indicators.

Simulation Results

The bias, RMSE, and the 95% coverage rate of the TIE are given in Table 2. The bias and RMSE of TIE were calculated as follows.

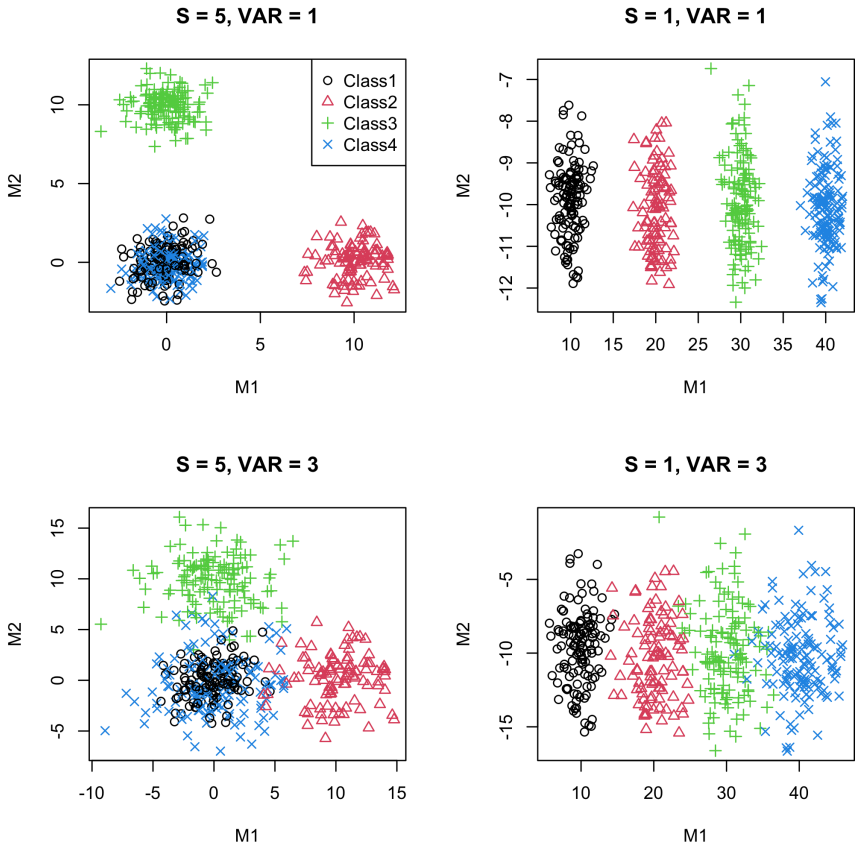
$$\begin{aligned} Bias &= \sum_{r=1}^{R} \frac{\widehat{TIE}_r - TIE}{R}, \\ RMSE &= \sqrt{\sum_{r=1}^{R} \frac{(\widehat{TIE}_r - TIE)^2}{R}}. \end{aligned} \tag{24}$$

$\widehat{TIE}_r$ is the total indirect effect estimate calculated based on the model parameter estimates in the rth replication, and TIE is the true total indirect effect. The proposed

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Figure 4

Scatter Plots of Simulated Indicators from the Simulation Conditions



latent class mediation analysis procedure recovered the total indirect effect of the latent class mediator well, although the total indirect effect was slightly overestimated. The magnitude of the bias slightly increased in the $VAR = 3$ conditions where the latent classes were allowed to overlap. However, the bias of TIE is negligible as the relative biases were less than 0.1 except for conditions 4 and 5. Further, we found that the bias of model parameter estimates decreased as the sample size increased in an additional simulation (Table A1). The 95% coverage rate was computed using 95% confidence intervals constructed from standard error estimates derived via the delta method.

$$95\%\ C.R.(TIE) = \sum_{r=1}^{R} \frac{I_{TIE \in (\widehat{TIE}_{L,r}, \widehat{TIE}_{U,r})}}{R}. \tag{25}$$

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

$\widehat{TIE}_{L,r}$ and $\widehat{TIE}_{U,r}$ are the lower bound and the upper bound of the 95% confidence interval, and I is the indicator function. The coverage rates of TIE were acceptable, ranging from .90 to .94 in the non-overlapping classes conditions and from .74 to .93 in the overlapping classes conditions.

Table 2

Simulation Study Results

Con.	VAR	S	ARI	N.Class	N.Ind	FP	TP	TIE		
								Bias	RMSE	95% C.R.
1	1	5	0.91	3.69	3.20	0.06	0.58	0.006	0.024	0.92
2	1	3	0.92	3.72	3.39	0.10	0.91	0.006	0.026	0.90
3	1	1	0.96	4.32	2.37	0.15	1.00	0.003	0.024	0.94
4	3	5	0.80	3.50	3.31	0.07	0.60	0.014	0.027	0.88
5	3	3	0.82	3.65	3.48	0.13	0.86	0.017	0.029	0.74
6	3	1	0.88	4.11	1.89	0.10	1.00	0.003	0.025	0.93

Note. 95% C.R. is the 95% coverage rate.

Throughout the simulation conditions, the selected number of classes was close to the true number of classes, $L = 4$, ranging from 3.50 to 4.32 (Table 2). The classification accuracy of the proposed analysis was evaluated using the average adjusted Rand index (ARI; Hubert and Arabie, 1985) between the estimated class and the true class. ARI measures the agreement of the two classifications when the number of classes does not necessarily match. ARI close to 1 indicates perfect agreement with the true classification, and ARI close to 0 indicates random classification. The formula of the ARI is given as follows. Let n_{ij} be the number of individuals in class i classified into the jth class. $L = 4$ is the number of true classes, and $\hat{L}$ is the number of classes in the latent class mediation model. Then,

$$ARI = \frac{\sum_{i=1}^{L}\sum_{j=1}^{\hat{L}}\binom{n_{ij}}{2} - [\sum_i \binom{n_{i.}}{2}\sum_j \binom{n_{.j}}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{n_{i.}}{2}\sum_j \binom{n_{.j}}{2}] - [\sum_i \binom{n_{i.}}{2}\sum_j \binom{n_{.j}}{2}]/\binom{n}{2}}, \tag{26}$$

where $n_{i.} = \sum_j^{\hat{L}} n_{ij}$, $n_{.j} = \sum_i^{L} n_{ij}$, and $n = \sum_i^{L}\sum_j^{\hat{L}} n_{ij}$. In the simulation conditions, the average ARI values were greater than .8, indicating an accurate classification of the proposed analysis. The ARI values were greater in the non-overlapping ($VAR = 1$)

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

condition (.91 $\sim$ .96) than the overlapping ($VAR = 3$) condition (.80 $\sim$ .88).

The variable selection algorithm performed well under the simulation conditions. In Table 2, the sixth column shows the average number of indicators selected in each condition. When three signal indicators were needed to identify the four true latent classes (i.e., conditions 1, 2, 4, and 5), slightly more than three variables were selected. When the first indicator was the only signal indicator (i.e., conditions 3 and 6), 2.37 and 1.89 indicators were selected on average in the final model. The seventh and eighth columns in Table 2 show the false positive rate and the true positive rate of selecting the indicator. The false positive rate is calculated as the probability of selecting a noisy indicator, and the true positive rate is calculated as the probability of selecting a signal indicator.

$$\begin{aligned} FP &= \sum_{r=1}^{R} \sum_{j=S+1}^{K} \frac{I_{M_j \in \mathbf{M}^r_{\kappa^*}}}{R(K-S)}. \\ TP &= \sum_{r=1}^{R} \sum_{j=1}^{S} \frac{I_{M_j \in \mathbf{M}^r_{\kappa^*}}}{RS}. \end{aligned} \tag{27}$$

$\mathbf{M}^r_{\kappa^*}$ is the set of indicators selected in the final model in the rth replication. The variable selection algorithm controlled the false positive rate reasonably, ranging from .06 to .15. In the $S = 5$ conditions, the true positive rate was .58 and .60, which means about 60% of the first five signal variables were selected, which suffices to identify the four true latent classes. In the $S = 3$ conditions, most of the three signal indicators were selected with the true positive rates of .91 and .86. In the $S = 1$ condition, the sole signal indicator was always selected in the final model with 1.00 true positive rate.

We conducted additional simulations to evaluate the accuracy of parameter estimates given the true number of latent classes, L. The EM algorithm for the LCMA model performed well, exhibiting low bias and RMSE in the parameter estimates. Additionally, we assessed the proposed algorithm's performance under alternative data-generating models. The algorithm showed robust performance across various scenarios in terms of both variable selection and parameter estimation. Further details about the simulation methods and results are provided in Appendices D and E.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

NAEP Math Assessment Data Analysis Results

The latent class mediation analysis was applied to the empirical data from the motivating example. To start, we fit a simple logistic regression predicting the final outcome $Y \in \{0, 1\}$ with the disability group membership G, without any mediator. The log odds of correct response were 1.25 lower in the LD group than in the TD group,

$$logitP(Y = 1|G) = -0.05 - 1.25G. \tag{28}$$

Without any mediators, the total effect of the group membership on the final outcome was -0.273, calculated as follows:

$$TE = E(Y|G = 1) - E(Y|G = 0). \tag{29}$$

Then, the proposed LCMA procedure is applied to the empirical data. Specifically, in the current context, the latent class mediation analysis aims to address the following research questions (RQs):

RQ1 What are the latent classes (Ω) of action sequence patterns that explain the relationship between disability group (G) and outcome (Y)? In other words, we search for Ω underlying M in Equations 3-4.

RQ2 What subset of action sequence features ($\mathbf{M}_\kappa, \kappa \subseteq \{1, \ldots, K\}$) can best account for the effect of disability group on the outcome? In other words, we search for $\kappa^* = \arg\max_\kappa TIE$ in Equations 22-23.

RQ3 How much of the group difference in final outcome can be explained by the latent class mediator (Ω) underlying problem-solving process features? In other words, we estimate and evaluate TIE in Equations 22-23.

The headlong search algorithm described previously was implemented to find the subset of indicators maximizing the total indirect effect of the disability group membership on the final score through the process features. Out of the $K = 15$ MDS process features, the

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

variable selection algorithm selected 14 indicators in the final model. The data analysis required approximately 31 hours with a sample size of N = 3090, K = 15 candidate features, and the maximum number of latent classes set to L = 20. The selected number of latent classes was $L = 18$. After incorporating the latent class mediator, the total indirect effect estimate was $\widehat{TIE} = -0.154$, controlling for the ELL variable, with a 95% confidence interval of $(-0.183, -0.125)$. This shows that the latent class variable underlying the selected process features could substantially explain the final score difference between LD and TD students, controlling for the ELL status to 0. To evaluate the reproducibility of our results, we randomly sampled 80% of the data four times and applied the proposed LCMA procedure to each subsample, examining the stability of both the TIE estimates and classification of students across the subsamples. Each subsample consisted of 470 LD students and 2000 TD students. The TIE estimates varied only slightly, from -0.119 to -0.157, across the four subsamples. While the optimal number of classes was 20 in these subsamples, the ARIs comparing classification from the total sample to those from the subsamples ranged from 0.963 to 0.979 indicating high consistency in classification of students.

To interpret and label the identified latent classes, we propose inspecting common patterns in the original action sequences of test takers within each class. Although a common approach involves describing classes based on their indicators (Spurk et al., 2020), this can be challenging with MDS features, as the extracted feature values are often difficult to interpret. Analyzing the action sequence offers a clearer and more practical approach to understanding and labeling the latent classes underlying the process data. Table 3 presents a descriptive summary of common patterns in the original sequences for each class, along with their corresponding class labels. Marked in (h) are homogeneous classes with identical action sequences. For instance, the common action sequence for Class 2, labeled “Revisit for review, 735”, was (Part_1_735, Exit_Item, Enter_Item). This indicates that every student in this class entered an answer with the numbers 7, 3, and 5

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

and revisited the item page once. On the other hand, the common action sequence for Class 5, labeled "Omission of the first try," was (Exit_Item, Enter_Item, Part_1_735). This indicates that every student in this class initially omitted the item and then submitted an answer with the numbers 7, 3, and 5 during their second visit to the item page. To interpret the non-homogeneous classes, we examined both the common actions within each class and the summary statistics provided in Table 3. For example, Class 1 was labeled as "Multiple revisits, no tools, 735", and every student in this class revisited the item page multiple times while submitting an answer with the numbers 7, 3, and 5.

Table 4 shows the model-implied correct response probabilities, $P(Y = 1)$, and class probabilities, $P(\Omega = l)$, for LD and TD students, along with their (log) odds ratios and the raw differences in class probabilities. These probabilities are calculated based on the model parameter estimates related to the logistic regression in Equation 4, γ and $\boldsymbol{\alpha}$, and the multinomial logistic regression in Equation 3, $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$, controlling for the covariate. Note that after classifying students into the latent classes, the difference in the correct response probabilities within each class between LD and TD students has decreased. This also shows that the latent class variable can explain the performance gap between the two groups. Behaviors observed in classes 1 to 9 are associated with higher correct response probabilities compared to the marginal correct response probability, while behaviors common in classes 10 to 18 are associated with lower correct response probabilities.

From Table 4, we identify the test-taking behaviors that contribute to the performance gaps between LD and TD students by focusing on the latent classes with substantial class probability $P(\Omega = l)$ differences in both the odds ratio and absolute difference scales. Since most class probabilities are small except for Classes 6 and 16, some absolute proportion differences are also small. The classes with higher correct response probabilities were Class 2 "Revisit for review, 735", Class 4 "Draw_Clear", Class 6 "No tools, 735", and Class 7 "Single draw". The class probability odds ratios for these classes were $0.33, 0.29, 0.44$, and 0.69, indicating that LD students were less likely to belong to

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Table 3

Tool Usage Rates of Latent Classes from the NAEP Math Assessment from the NAEP Math Assessment Item VH336968

Class	Label	No.	Len.M	Len.SD	Rev.	Avg.Rev.	Dra.	Era.	Cle.	E | C	Hig.	Zoo.	The.	TOS
1	Multiple revisits, no tools, 735	40	5.16	0.69	1.00	2.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	Revisit for review, 735 (h)	150	3.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	Draw_Erase	90	6.28	2.63	0.03	0.04	1.00	1.00	0.42	1.00	0.03	0.00	0.00	0.00
4	Draw_Clear	60	4.95	1.53	0.00	0.00	1.00	0.00	0.98	0.98	0.00	0.00	0.00	0.00
5	Omission of the first try (h)	50	3.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	No tools, 735 (h)	1020	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	Single draw (h)	140	2.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	Draw and revisit	100	6.61	2.20	1.00	1.17	1.00	0.43	0.57	0.79	0.01	0.00	0.00	0.00
9	Draw with clear or erase, revisit	170	7.53	5.96	0.41	0.74	0.90	0.49	0.56	0.83	0.11	0.01	0.05	0.08
10	Irrelevant tools (TOS) or reentries	80	5.72	2.48	0.31	0.37	0.52	0.16	0.17	0.29	0.03	0.00	0.03	0.91
11	Irrelevant tools (theme) or revisit	70	6.40	3.38	0.30	0.40	0.44	0.14	0.19	0.29	0.01	0.10	0.81	0.14
12	Draw with clear or erase	130	6.53	3.31	0.02	0.03	1.00	0.66	0.63	0.95	0.02	0.00	0.00	0.01
13	Multiple revisits or reentries	60	6.63	2.94	1.00	2.32	0.11	0.05	0.08	0.10	0.00	0.00	0.00	0.02
14	Omission of the first try, non-735 (h)	40	3.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	Draw_Erase or Draw_Clear, non-735	30	3.00	0.00	0.00	0.00	1.00	0.41	0.59	1.00	0.00	0.00	0.00	0.00
16	No tools, non-735 (h)	720	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	Revisit for review, non-735 (h)	60	3.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	Single draw, non-735 (h)	70	2.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note. No.: Number of students classified into each class (rounded to the nearest ten); Len.M: Average action sequence length; Len.SD: Standard deviation of action sequence length; Rev. Revisit; Avg.Rev.: Average number of revisits; Dra.: Draw; Era.: Erase; Cle.: Clear; E | C: Erase or Clear; Hig.: Highlight; Zoo.: Zooming in/out; The.: Theme editor; TOS.: text-to-speech;

SOURCE: U.S. Department of Education, National Center for Education Statistics, "Response Process Data from the NAEP 2017 Grade 8 Mathematics Assessment."

Table 4

Model Implied Response Probabilities and Class Probabilities from the NAEP Math Assessment Item VH336968

Class	Label	$P(Y=1)$		$P(\Omega=l)$				
		LD	TD	LD	TD	OR	LOR	DIFF
	Marginal probability	0.21	0.49					
1	Multiple revisits, no tools, 735	0.75	0.87	0.00	0.02	0.21	-0.67	-0.01
2	Revisit for review, 735 (h)	0.74	0.86	0.02	0.05	0.33	-0.48	-0.04
3	Draw_Erase	0.63	0.79	0.03	0.03	0.82	-0.09	-0.01
4	Draw_Clear	0.60	0.77	0.01	0.02	0.29	-0.53	-0.02
5	Omission of the first try (h)	0.60	0.77	0.01	0.02	0.44	-0.35	-0.01
6	No tools, 735 (h)	0.59	0.75	0.20	0.36	0.44	-0.36	-0.16
7	Single draw (h)	0.57	0.74	0.04	0.05	0.69	-0.16	-0.02
8	Draw and revisit	0.35	0.54	0.03	0.03	1.07	0.03	0.00
9	Draw with clear or erase, revisit	0.34	0.53	0.06	0.06	1.04	0.02	0.00
10	Irrelevant tools (TOS) or reentries	0.20	0.35	0.03	0.02	1.59	0.20	0.01
11	Irrelevant tools (theme) or revisit	0.19	0.34	0.03	0.02	1.41	0.15	0.01
12	Draw with clear or erase	0.01	0.02	0.05	0.04	1.34	0.13	0.01
13	Multiple revisits or reentries	0.00	0.00	0.03	0.01	2.46	0.39	0.02
14	Omission of the first try, non-735 (h)	0.00	0.00	0.02	0.01	1.63	0.21	0.01
15	Draw_Erase or Draw_Clear, non-735	0.00	0.00	0.02	0.00	5.31	0.73	0.02
16	No tools, non-735 (h)	0.00	0.00	0.36	0.20	2.20	0.34	0.16
17	Revisit for review, non-735 (h)	0.00	0.00	0.03	0.02	1.59	0.20	0.01
18	Single draw, non-735 (h)	0.00	0.00	0.03	0.02	1.40	0.15	0.01

Note. $P(Y=1)$ is the model implied correct response probability. $P(\Omega=l)$ is the model implied probability of belonging to the l-th class. (h) indicates a homogeneous class with the same action sequence. OR: Model implied odds ratio of class probabilities for LD against TD; LOR: Log odds ratio; DIFF: Difference in class probabilities.

SOURCE: U.S. Department of Education, National Center for Education Statistics, "Response Process Data from the NAEP 2017 Grade 8 Mathematics Assessment."

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

these classes. Specifically, LD students were less likely to revisit the item for review and submit an answer with numbers 7, 3, and 5. Additionally, behaviors such as using scratchwork with a single draw stroke or clearing the scratchwork immediately after drawing led to higher correct response probabilities, yet LD students were less likely to display these behaviors. When using no tools, LD students were less likely to submit an answer containing 7, 3, and 5, suggesting they were more likely to make non-decimal point errors and demonstrate misconceptions in their responses. On the other hand, LD students were more likely to belong to the low-performing classes, Class 13 "Multiple revisits or reentries", Class 15 "Draw_Erase or Draw_Clear, non-735", and Class 16 "No tools, non-735". The class probability odds ratios were $2.46, 5.31$, and 2.20, respectively. These results suggest that the behaviors associated with worse performance, and more commonly observed among LD students, include multiple revisits, a sequence of Draw and Erase or Clear with non-735 responses, and using no tools with non-735 responses.

These results show key differences in test-taking behaviors between LD and TD students, particularly in their use of scratchwork, item review, and response patterns for non-735 answers. TD students were more likely to engage in effective scratchwork strategies, such as making a single draw stroke, which were associated with higher correct response probabilities. In contrast, LD students tended to engage in unproductive behaviors like repeatedly revisiting or re-entering answers, which are associated with lower performance. Additionally, for students who submitted a non-735 answer, common incorrect responses such as $4.45, 8.5$, and 29.4 suggest deeper misconceptions about decimal multiplication. These answers could be derived by the following computations: $(4 \times 1) + (0.9 \times 0.5) = 4.45$; $(4 \times 1) + (9 \times 0.5) = 8.5$; and $(4.9 \times 1) + (4.9 \times 5) = 29.4$. Each of these errors goes beyond simple misplacement of the decimal point and indicates fundamental misunderstandings about multiplication rules in the context of decimals.

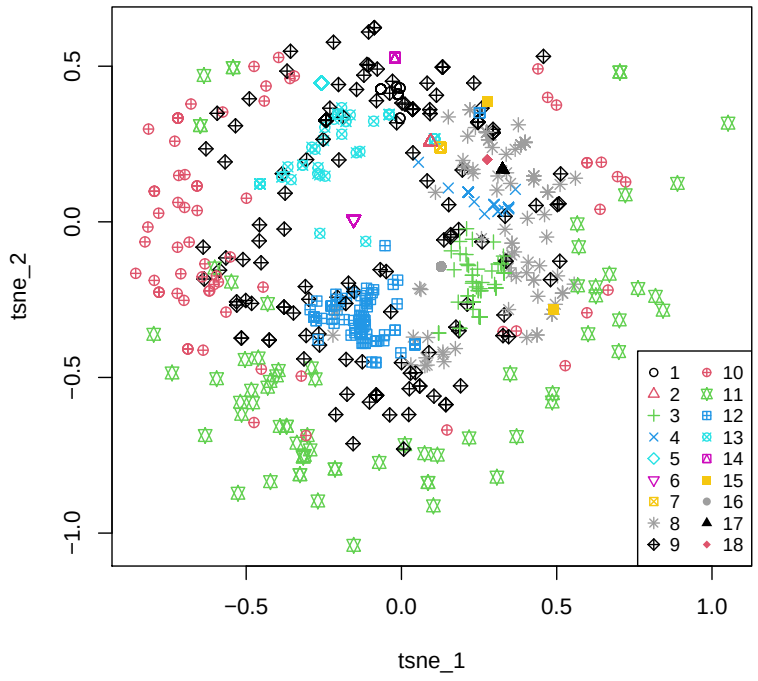
In Figure 5, the global structure of the selected process features is displayed on a two-dimensional plot using the t-distributed stochastic neighborhood embedding (t-SNE;

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Van der Maaten and Hinton, 2008). The t-SNE is a popular dimension-reduction method for visualization that preserves the similarity between observations by considering the observation's nearest neighbors. While Figure 5 displays a grouping of the 18 classes into distinct areas, some classes are less clearly separated. The homogeneous classes are displayed as single points, as these classes had identical feature values. Classes 11 and 10 were the most dispersed classes in the plot because students in these classes were randomly browsing the available tools and submitted various responses.

Figure 5

t-SNE Plot of the Selected Process Features from the NAEP Math Assessment Item VH336968



SOURCE: U.S. Department of Education, National Center for Education Statistics, "Response Process Data from the NAEP 2017 Grade 8 Mathematics Assessment."

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Discussion

Process data collected in computerized testing preserves valuable information beyond the traditional response data. However, analyzing process data is challenging because of its unstructured format and noise, which hinders the use of traditional approaches developed for rectangular data. This study provides an approach to a traditionally challenging task with new but noisy process data. The proposed LCMA analysis procedure is a general statistical method that can be applied when the latent class variable underlying action sequences is assumed as the mediator between an observed predictor and an outcome. The latent class mediation model and the headlong search algorithm allow dimension reduction and noise elimination from the process features, enhancing the interpretability of the results.

The latent class analysis with continuous indicators, often called latent profile analysis or Gaussian mixture clustering, is extended to a latent class mediation analysis. To the best of our knowledge, the current study is the first attempt to extend the latent class analysis assuming multivariate normality of the indicators into a latent class mediation model including both a covariate and a distill outcome to assess the mediation effect of the nominal latent class variable. There are few studies using a latent class mediator with continuous indicators. For example, Sint et al. (2021) proposed a latent class mediation analysis where the observed continuous indicator was specified as a generalized linear model, given the latent class. The limitation of such an approach is that the covariance structure of the indicators was not considered.

Process data from large-scale assessments can help understand why certain students are struggling, serving as a seminal guide to efforts on evidence-based strategies to improve educational equity. The proposed analysis can help educators design targeted treatments for specific subgroups. With the NAEP Math Assessment data, we showed that the proposed latent class mediation analysis can identify the latent class variable that explains the performance gap on a multiplication item between the students with learning disability

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

and the typically developing students. Each class was interpreted and labeled based on summary statistics, such as the tool usage rates of the students classified into each class. Then, calculating the model-implied correct response probabilities and class probabilities using the parameter estimates from the proposed model allowed us to attribute the performance gap between the two groups to the difference in test-taking behaviors. The key point is that identifying the latent classes underlying the features and examining how the two groups differ in their probabilities of belonging to each latent class allows us to explain the performance gap between the two groups.

Practical implications of the NAEP Math Assessment data analysis demonstrate the importance of identifying specific test-taking behaviors that led to performance gaps between LD and TD students. By focusing on behaviors such as revisiting questions and employing effective problem-solving strategies, educators can design targeted interventions to help LD students develop more effective test-taking habits and improve their overall performance. Additionally, grouping students based on their specific test-taking behaviors can allow teachers to provide more focused support and instruction to meet individual needs, and such strategies can help bridge the gap in academic performance between LD and TD students.

The current study implemented a simultaneous estimation method for the latent class mediation model using an EM algorithm. The proposed estimation method is justified by the simulation results as the model parameters were accurately estimated when the data was generated from the true model. In addition, Bolck et al. (2004) suggested that simultaneous estimation is viable in latent class analysis with continuous indicators when a distal outcome is predicted by the latent class variable. However, in the mediation analysis with a latent variable, variations of two-step and three-step estimation approaches with adjustments for classification errors may be available. In the context of latent class mediation analysis with categorical indicators, Hsiao et al. (2021) compared six different estimation methods, including variations of one-step, two-step, and three-step approaches.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

There is a demand for an investigation of the estimation in the latent class mediation analysis with continuous indicators in various conditions.

A headlong search algorithm for feature selection is proposed. The objective of the feature selection algorithm is to find the subset of process features that maximizes the total indirect effect of the latent class mediator. In the simulation, the proposed feature selection algorithm performed well in selecting the signal features while excluding the noisy features irrelevant to the true clustering. This approach aligns with the idea of the exploratory mediation analysis (van Kesteren & Oberski, 2019) where a mediation filter was used to find the subset of many potential mediators to explain the effect of the predictor on the outcome. There is one caveat to the proposed feature selection algorithm. Each inclusion and exclusion step requires significance testing. As the number of iterations in the search algorithm increases, the family-wise type-1 error can be hard to control at a desired significance level. Therefore, family-wise type-1 error control methods proposed in step-wise variable selection, such as Bonferroni correction, may be considered. Or, considering different criteria, such as a decrease in the direct effect for selecting the initial set of features, may improve the reliability of the search algorithm. Another alternative could be implementing a search algorithm that does not rely on step-wise decisions.

We adopted the counterfactual approach (Pearl, 2010; Robins & Greenland, 1992) and the formal definitions of effects involving a latent class mediator described in Muthén (2011) to assess the total indirect effect of the nominal latent class mediator. The indirect effects defined in the counterfactual framework rely on several strict assumptions and are described in Imai, Keele, and Tingley (2010), Valeri and VanderWeele (2013), and VanderWeele and Vansteelandt (2009). A part of the assumptions can be satisfied when the predictor is a randomized treatment. Other assumptions require that there is no unmeasured confounding variable of the predictor-outcome relationship and the mediator-outcome relationship. The effects of unmeasured confounding variables can be controlled by including them as covariates, as described previously. In observational

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

research, however, including demographic variables such as learning disability status may still violate the randomized treatment assumption. The indirect effect estimates are biased when some of the assumptions are violated.

Importantly, we emphasize that the proposed framework is intended as an exploratory tool for generating hypotheses about causal relationships in complex process data, rather than for drawing causal claims about test-taking behavior. To advance from hypothesis generation to more robust causal statements, future work could integrate formal sensitivity analyses. For example, future studies could adopt bias-adjustment formulas for unmeasured confounding (VanderWeele & Arah, 2011), sensitivity analysis for causal mediation effects (Imai, Keele, & Yamamoto, 2010), or statistical methods for examining and adjusting for assumption violations (MacKinnon & Pirlott, 2015).

Complex latent-class models are susceptible to convergence at local optima, which can in turn affect BIC-based model selection. We initialize the EM algorithm via hierarchical agglomeration clustering, as implemented in the `mclust` R package (Scrucca et al., 2023), to optimize the chance of arriving at an accurate model solution. Nonetheless, future extensions could consider incorporating multiple-start EM runs, as implemented in `Mplus` (Muthén & Muthén, 2017). It should also be noted that uncertainty in the BIC-based model selection could propagate to mediation effect estimates. Such unaddressed model selection variability may lead to underestimation of posterior uncertainty for indirect effect parameters. To address this, one can adopt fully Bayesian model approaches treating the number of latent classes as a random variable (see e.g., Chen et al., 2021; Richardson and Green, 1997; Stephens, 2000) or apply Bayesian model averaging over candidate models (see e.g., Hoeting et al., 1999; Russell et al., 2015; Wasserman, 2000).

Other machine learning techniques can be used to extract process features from the unstructured action sequence data while preserving the information of the original data. The type of information kept in the process features will depend on the feature extraction

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

method. For example, N-gram-based techniques could extract the frequencies of a sequence of actions (e.g., He and von Davier, 2016). One potential advantage of using N-gram-based features in latent class mediation modeling is that the selected features can be more interpretable. Each feature is related to the frequency of a certain action sequence. Therefore, the selected features can directly show the test-taking behavior that explains performance gaps between groups. However, the N-gram features are discrete variables, and the multivariate normality assumption of the proposed analysis may not hold. A future extension of this study could involve incorporating discrete features, such as count or binary data, into the proposed analysis framework. Another possible direction is extending the model to accommodate a multi-categorical group membership predictor by introducing $C - 1$ dummy variables with their corresponding regression coefficients, where C represents the number of categories.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

References

Banfield, J. D., & Raftery, A. E. (1993). Model-based gaussian and non-gaussian clustering. *Biometrics*, 803–821.

Bergner, Y., & von Davier, A. A. (2019). Process data in naep: Past, present, and future. *Journal of Educational and Behavioral Statistics*, *44*(6), 706–732.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political analysis*, *12*(1), 3–27.

Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, *6*(1), 76–90.

Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern recognition*, *28*(5), 781–793.

Chen, Y., Liu, Y., Culpepper, S. A., & Chen, Y. (2021). Inferring the number of attributes for the exploratory dina model. *Psychometrika*, *86*(1), 30–64.

Chen, Y., et al. (2020). A continuous-time dynamic choice measurement model for problem-solving process data. *Psychometrika*, *85*(4), 1052–1075.

Dean, N., & Raftery, A. E. (2010). Latent class analysis variable selection. *Annals of the Institute of Statistical Mathematics*, *62*, 11–35.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, *39*(1), 1–22.

Dziak, J. J., Bray, B. C., Zhang, J., Zhang, M., & Lanza, S. T. (2016). Comparing the performance of improved classify-analyze approaches for distal outcomes in latent profile analysis. *Methodology*.

Eichmann, B., Goldhammer, F., Greiff, S., Brandhuber, L., & Naumann, J. (2020). Using process data to explain group differences in complex problem solving. *Journal of educational psychology*, *112*(8), 1546.

Eichmann, B., Greiff, S., Naumann, J., Brandhuber, L., & Goldhammer, F. (2020). Exploring behavioural patterns during complex problem-solving. *Journal of Computer Assisted Learning*, *36*(6), 933–956.

Fang, G., Ying, Z., et al. (2020). Latent theme dictionary model for finding co-occurrent patterns in process data. *Psychometrika*, *85*(3), 775–811.

Fletcher, R. (1970). A new approach to variable metric algorithms. *The computer journal*, *13*(3), 317–322.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*(458), 611–631.

Gao, Y., Cui, Y., Bulut, O., Zhai, X., & Chen, F. (2022). Examining adults' web navigation patterns in multi-layered hypertext environments. *Computers in Human Behavior*, *129*, 107142.

Gao, Y., Zhai, X., Bulut, O., Cui, Y., & Sun, X. (2022). Examining humans' problem-solving styles in technology-rich environments using log file data. *Journal of Intelligence*, *10*(3), 38.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of computation*, *24*(109), 23–26.

Gómez-Alonso, C., & Valls, A. (2008). A similarity measure for sequences of categorical data based on the ordering of common elements. *Modeling Decisions for Artificial Intelligence: 5th International Conference, MDAI 2008 Sabadell, Spain, October 30-31, 2008. Proceedings 5*, 134–145.

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? a showcase study based on the pisa 2012 assessment of problem solving. *Computers & Education*, *91*, 92–105.

Hao, J., & Mislevy, R. J. (2019). Characterizing interactive communications in computer-supported collaborative problem-solving tasks: A conditional transition profile approach. *Frontiers in Psychology*, *10*, 424340.

He, Q., Borgonovi, F., & Suárez-Álvarez, J. (2022). Clustering sequential navigation patterns in multiple-source reading tasks with dynamic time warping method. *Journal of computer-assisted learning*, 719–736. https://doi.org/10.1111/jcal.12748

He, Q., Borgonovi, F., & Suárez-Álvarez, J. (2023). Clustering sequential navigation patterns in multiple-source reading tasks with dynamic time warping method. *Journal of Computer Assisted Learning*, *39*(3), 719–736.

He, Q., Shi, Q., & Tighe, E. L. (2023). Predicting problem-solving proficiency with multiclass hierarchical classification on process data: A machine learning approach. *Psychological Test and Assessment Modeling*, *65*(1), 145–177.

He, Q., & von Davier, M. (2016). Analyzing process data from problem-solving items with n-grams: Insights from a computer-based large-scale assessment. In *Handbook of research on technology tools for real-world skill development* (pp. 750–777). IGI Global.

He, Q., von Davier, M., & Han, Z. (2019). Exploring process data in problem-solving items in computer-based large-scale assessments: Case studies in PISA and PIAAC. In H. Jiao, R. W. Lissitz, & A. Van Wie (Eds.), *Data analytics and psychometrics: Informing assessment practices*. Information Age Publishing, Inc. Charlotte, NC.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. *Statistical science*, *14*(4), 382–417.

Hsiao, Y.-Y., Kruger, E. S., Lee Van Horn, M., Tofighi, D., MacKinnon, D. P., & Witkiewitz, K. (2021). Latent class mediation: A comparison of six approaches. *Multivariate Behavioral Research*, *56*(4), 543–557.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, *2*, 193–218.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, *15*(4), 309.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71.

Judge, S., & Watson, S. M. (2011). Longitudinal outcomes for mathematics achievement for students with learning disabilities. *The Journal of Educational Research*, *104*(3), 147–157.

Keribin, C. (1998). Consistent estimate of the order of mixture models. *Comptes Rendus De L Academie Des Sciences Serie I-Mathematique*, *326*(2), 243–248.

LaMar, M. M. (2018). Markov decision process measurement model. *Psychometrika*, *83*(1), 67–88.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. *Studies in social psychology in world war II Vol. IV: Measurement and prediction*, 362–412.

Lazarsfeld, P. F. (1968). Latent structure analysis. *(No Title)*.

Liao, D., He, Q., & Jiao, H. (2019). Mapping background variables with sequential patterns in problem-solving environments: An investigation of united states adults' employment status in piaac. *Frontiers in Psychology*, *10*, 646.

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the m to y relation in mediation analysis. *Personality and Social Psychology Review*, *19*(1), 30–43.

Murphy, K., & Murphy, T. B. (2020). Gaussian parsimonious clustering models with covariates and a noise component. *Advances in Data Analysis and Classification*, *14*(2), 293–325.

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of classification*, *31*, 274–295.

Muthén, B. (2011). Applications of causally defined direct and indirect effects in mediation analysis using sem in mplus.

Muthén, B., & Muthén, L. (2017). Mplus. In *Handbook of item response theory* (pp. 507–518). Chapman; Hall/CRC.

NCES. (2020). Process data from the 2017 naep grade 8 mathematics assessment [Accessed: June 15, 2024].

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. *Modern statistical methods for HCI*, 275–287.

Pearl, J. (2010). An introduction to causal inference. *The international journal of biostatistics*, *6*(2).

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual review of psychology*, *66*(1), 825–852.

Qiao, X., & Jiao, H. (2018). Data mining techniques in analyzing process data: A didactic. *Frontiers in psychology*, 2231.

Richardson, S., & Green, P. J. (1997). On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *59*(4), 731–792.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400–407.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*(2), 143–155.

Russell, N., Murphy, T. B., & Raftery, A. E. (2015). Bayesian model averaging in model-based clustering and density estimation. *arXiv preprint arXiv:1506.09035*.

Scrucca, L., Fraley, C., Murphy, T. B., & Raftery, A. E. (2023). *Model-based clustering, classification, and density estimation using mclust in R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781003277965

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, *24*(111), 647–656.

Sint, K., Rosenheck, R., & Lin, H. (2021). Latent class mediator for multiple indicators of mediation. *Statistics in medicine*, *40*(12), 2800–2820.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of vocational behavior*, *120*, 103445.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of statistics*, 40–74.

Takane, Y. (2006). 11 applications of multidimensional scaling in psychometrics. *Handbook of statistics*, *26*, 359–400.

Tang, X. (2023). A latent hidden markov model for process data. *Psychometrika*, 1–36.

Tang, X., Wang, Z., He, Q., Liu, J., & Ying, Z. (2020). Latent feature extraction for process data via multidimensional scaling. *Psychometrika*, *85*, 378–397.

Tang, X., Zhang, S., Wang, Z., Liu, J., & Ying, Z. (2021). Procdata: An r package for process data analysis. *psychometrika*, *86*(4), 1058–1083.

Ulitzsch, E., He, Q., & Pohl, S. (2022). Using sequence mining techniques for understanding incorrect behavioral patterns on interactive tasks. *Journal of Educational and Behavioral Statistics*, *47*(1), 3–35.

Ulitzsch, E., He, Q., Ulitzsch, V., Molter, H., Nichterlein, A., Niedermeier, R., & Pohl, S. (2021). Combining clickstream analyses and graph-modeled data clustering for identifying common response processes. *Psychometrika*, *86*(1), 190–214.

Ulitzsch, E., Ulitzsch, V., He, Q., & Lüdtke, O. (2022). A machine learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks. *Behavior Research Methods*, 1–21.

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2022). National assessment of educational progress (naep), 2022 reading assessment [Accessed: June 15, 2024].

Valente, M. J., Pelham III, W. E., Smyth, H., & MacKinnon, D. P. (2017). Confounding in statistical mediation analysis: What it is and how to address it. *Journal of counseling psychology*, *64*(6), 659–671.

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with sas and spss macros. *Psychological methods*, *18*(2), 137.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, *9*(11).

VanderWeele, T. J., & Arah, O. A. (2011). Unmeasured confounding for general outcomes, treatments, and confounders: Bias formulas for sensitivity analysis. *Epidemiology (Cambridge, Mass.)*, *22*(1), 42–52.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, *2*(4), 457–468.

van Kesteren, E.-J., & Oberski, D. L. (2019). Exploratory mediation analysis with many potential mediators. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 710–723.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political analysis*, *18*(4), 450–469.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied latent class analysis*, *11*(89-106), 60.

Wang, Z., Tang, X., Liu, J., & Ying, Z. (2020). Subtask analysis of process data through a predictive model. *arXiv preprint arXiv:2009.00717*.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of mathematical psychology*, *44*(1), 92–107.

Welling, J., Gnambs, T., & Carstensen, C. H. (2024). Identifying disengaged responding in multiple-choice items: Extending a latent class item response model with novel process data indicators. *Educational and Psychological Measurement*, *84*(2), 314–339.

Witkiewitz, K., Roos, C. R., Tofighi, D., & Van Horn, M. L. (2018). Broad coping repertoire mediates the effect of the combined behavioral intervention on alcohol outcomes in the combine study: An application of latent class mediation. *Journal of studies on alcohol and drugs*, *79*(2), 199–207.

Xiao, Y., & Liu, H. (2024). A state response measurement model for problem-solving process data. *Behavior Research Methods*, *56*(1), 258–277.

Zhan, P., & Qiao, X. (2022). Diagnostic classification analysis of problem-solving competence using process data: An item expansion method. *Psychometrika*, *87*(4), 1529–1547.

Zhang, S., Wang, Z., Qi, J., Liu, J., & Ying, Z. (2022). Accurate assessment via process data. *Psychometrika*, 1–22.

Appendix A. Multidimensional Scaling for Action Sequence Data

Multidimensional scaling (Borg & Groenen, 2005) is a dimension reduction method that extracts latent features based on the pairwise dissimilarity measure between two observations. MDS is widely used for data visualization and in many areas of psychometrics (Takane, 2006). The goal of MDS is to locate observations within a vector space based on their pairwise dissimilarities, ensuring that similar observations are located closely. In contrast, less similar ones are located farther apart. Tang et al. (2020) proposed using multidimensional scaling for extracting process features from the problem-solving process data. In process data analysis, if dissimilarities effectively capture differences between two processes, the coordinates derived from MDS can serve as features containing information about the original processes (Tang et al., 2020).

The dissimilarity measure between two action sequences takes into account the number of unique actions and the order of common actions (Gómez-Alonso & Valls, 2008). Let $\mathbf{s}_i = (s_{i1}, ..., s_{iL_i})$ and $\mathbf{s}_j = (s_{j1}, ..., s_{jL_j})$ be two action sequences of examinee i and j. L_i and L_j are the lengths of each action sequence. C_{ij} denotes the set of common actions that appear in both $\mathbf{s}_i$ and $\mathbf{s}_j$. U_{ij} denotes the set of actions that appear in $\mathbf{s}_i$ but not in $\mathbf{s}_j$. Let $\mathbf{s}^a$ be the number of times that an action a appears in $\mathbf{s}$. $\mathbf{s}^a(k)$ denotes the kth element of $\mathbf{s}^a$ that is, the position of the kth appearance of a in $\mathbf{s}$. Then the dissimilarity among the common actions in $\mathbf{s}_i$ and $\mathbf{s}_j$ is quantified as

$$f(\mathbf{s}_i, \mathbf{s}_j) = \frac{\sum_{a \in C_{ij}} \sum_{k=1}^{K_{ij}^a} |\mathbf{s}_i^a(k) - \mathbf{s}_j^a(k)|}{\max\{L_i, L_j\}}, \tag{A1}$$

where $K_{ij}^a = \min(L_i^a, L_j^a)$. The count of unique actions appearing in only one of $\mathbf{s}_i$ and $\mathbf{s}_j$ is quantified as

$$g(\mathbf{s}_i, \mathbf{s}_j) = \sum_{a \in U_{ij}} L_i^a + \sum_{a \in U_{ji}} L_j^a. \tag{A2}$$

Then the dissimilarity between two action sequences is defined by

$$d(\mathbf{s}_i, \mathbf{s}_j) = \frac{f(\mathbf{s}_i, \mathbf{s}_j) + g(\mathbf{s}_i, \mathbf{s}_j)}{L_i + L_j}. \tag{A3}$$

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Let $D = (d_{ij})$ be the $N \times N$ symmetric dissimilarity matrix, where d_{ij} measures the dissimilarity between $\mathbf{s}_i$ and $\mathbf{s}_j$. Higher dissimilarities indicate greater disparities, and the dissimilarity between identical action sequences is zero. MDS assigns each action sequence to a latent vector $\mathbf{m}$ in the K-dimensional Euclidean space such that these vectors dictate the dissimilarities. The application of MDS to the dissimilarity matrix D minimizes

$$\sum_{i<j}(d_{ij} - ||\mathbf{m}_i - \mathbf{m}_j||)^2. \tag{A4}$$

The stochastic gradient descent (Robbins & Monro, 1951) can be used to solve the optimization problem. Let $\mathbf{M} = (\mathbf{m}_1, ..., \mathbf{m}_N)^T$ be the set of all process features extracted from the nation sequence process data. Then $\mathbf{M}$ has a standard form with homogeneous dimension while preserving the information of the original sequences. Hence, it can serve as a substitute for action sequences in traditional statistical models like generalized linear models (Tang et al., 2020). The number of process features K can be chosen by cross-validation and minimizing the loss function in Equation A4.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Appendix B. Approximation of the Standard Error of TIE by Delta Method

The TIE for the LCMA can be expressed as:

$$\begin{aligned} TIE &= E[Y(1,\Omega(1)) - Y(1,\Omega(0))] \\ &= \sum_{l=1}^{L} P(Y=1 \mid G=1, \Omega=l)\Big[P(\Omega=l \mid G=1) - P(\Omega=l \mid G=0)\Big] \\ &= \sum_{l=1}^{L} h(\alpha_l+\gamma)\Big[P(\Omega=l \mid G=1) - P(\Omega=l \mid G=0)\Big], \end{aligned} \tag{A5}$$

where

$$h(x) = \frac{e^x}{1+e^x}. \tag{A6}$$

Let's denote $P_{\omega|g}$ as,

$$P_{\omega|g} = P(\Omega=\omega \mid G=g) = \frac{exp(\beta_{0\omega}+\beta_{1\omega}g)}{\sum_{l=1}^{L} exp(\beta_{0l}+\beta_{1l}g)}. \tag{A7}$$

Then the partial derivative of $P_{\omega|g}$ with respect to β_{0l} and β_{1l} are

$$\frac{\partial P_{\omega|g}}{\partial \beta_{0l}} = D_l \text{ and } \frac{\partial P_{\omega|g}}{\partial \beta_{1l}} = gD_l, \quad \text{for } \omega = l, \tag{A8}$$

with

$$D_l = \frac{exp(\beta_{0l}+\beta_{1l}g)\left(\sum_{d=1,d\neq l}^{L} exp(\beta_{0d}+\beta_{1d}g)\right)}{\left(\sum_{d=1}^{L} exp(\beta_{0d}+\beta_{1d}g)\right)^2}, \tag{A9}$$

and

$$\frac{\partial P_{\omega|g}}{\partial \beta_{0l}} = -P_{\omega|g}P_{l|g} \text{ and } \frac{\partial P_{\omega|g}}{\partial \beta_{1l}} = -gP_{\omega|g}P_{l|g}, \quad \text{for } \omega \neq l. \tag{A10}$$

The partial derivatives of the TIE with respect to the parameters are:

$$\frac{\partial TIE}{\partial \beta_{il}} = \sum_{d=1}^{L} h(\alpha_d+\gamma)\left(\frac{\partial P_{d|1}}{\partial \beta_{il}} - \frac{\partial P_{d|0}}{\partial \beta_{il}}\right). \tag{A11}$$

$$\frac{\partial TIE}{\partial \alpha_l} = h'(\alpha_l+\gamma)\Big[P_{l|1} - P_{l|0}\Big], \tag{A12}$$

where

$$h'(x) = \frac{e^x}{(1+e^x)^2}. \tag{A13}$$

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

$$\frac{\partial TIE}{\partial \gamma} = \sum_{d=1}^{L} h'(\alpha_d + \gamma)\Big[P_{d|1} - P_{d|0}\Big]. \tag{A14}$$

The gradient of the TIE with respect to the parameters is:

$$\Gamma = \left(\frac{\partial TIE}{\partial \beta_{02}}, ..., \frac{\partial TIE}{\partial \beta_{0L}}, \frac{\partial TIE}{\partial \beta_{12}}, ..., \frac{\partial TIE}{\partial \beta_{1L}}, \frac{\partial TIE}{\partial \gamma}, \frac{\partial TIE}{\partial \alpha_1}, ..., \frac{\partial TIE}{\partial \alpha_L}\right) \tag{A15}$$

The approximation for the SE of the $\widehat{TIE}$ is then $\sqrt{\Gamma\Sigma\Gamma'}$, where Σ is the covariance matrix of the parameters. The $(1-\alpha)\%$ confidence interval of TIE is constructed as $\widehat{TIE} \pm z_{\alpha/2} \times SE$, where $z_{\alpha/2}$ is the critical value from the standard normal distribution.

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Appendix C. True Model Parameter Values in the Simulation Study

In the simulation study, the true model parameter values are set as follows. The true γ, $\boldsymbol{\alpha}$, $\boldsymbol{\beta_1}$, and $\boldsymbol{\beta_0}$ values were fixed in all simulation conditions.

$$\begin{aligned} \gamma &= 0, \\ \boldsymbol{\alpha} &= (-1, -1/3, 1/3, 1), \\ \boldsymbol{\beta_1} &= (0, -2/3, -4/3, -2), \\ \boldsymbol{\beta_0} &= (0, 1/3, 2/3, 1). \end{aligned} \tag{A16}$$

The true class-specific mean structure, $\underset{\sim}{\boldsymbol{\mu}}$ is given in Figure 3. The true class-specific covariance matrix is composed of $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_L)$ that controls the volume and a diagonal matrix B where the class-specific covariance matrix is $\Sigma_l = \lambda_l B$. In $Var = 1$ conditions, $\boldsymbol{\lambda}$ was set as

$$\boldsymbol{\lambda} = (1, 1, 1, 1), \tag{A17}$$

so that the volume of the four classes is equal, and small enough to yield no between-class overlap. In $Var = 3$ conditions, $\boldsymbol{\lambda}$ was set as

$$\boldsymbol{\lambda} = (3, 5, 7, 9). \tag{A18}$$

The classes were allowed to vary in their volumes and have overlapping observations between classes as demonstrated in Figure 4. The diagonal elements of B, $diag(B) = (B_{1,1}, ..., B_{K,K})$ were generated as follows. Let $B^* = (B_1^*, B_2^*, ..., B_K^*)$, where K is the number of items.

$$B_i^* = 1 + \frac{i-1}{45}, \quad i = 1, ..., K. \tag{A19}$$

The variance of the 10th item is 1.2 times the variance of the first item within a class. Then B^* was normalized by the geometric mean to satisfy $|B| = 1$.

$$B_{i,i} = \frac{B_i^*}{(\prod_{j=1}^{K} B_i^*)^{1/K}}, \quad i = 1, ..., K. \tag{A20}$$

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Appendix D. Model Parameter Recovery Check

In this section, we evaluate the accuracy of parameter estimates of the LCMA model. The LCMA model was fitted using all indicators $\mathbf{M}_\kappa$, with the number of latent classes L fixed at its true value, to assess the parameter estimation accuracy of the EM algorithm. Random samples were generated under a latent class mediation model with $L = 4$ latent classes and $K = 10$ signal indicators. The sample sizes considered were $N = 500$ and $N = 1000$. The true parameter values are specified as Equations A16-A17 and A19-A20. The additional parameters related to the covariate $\mathbf{X}$ were set as follows.

$$\begin{aligned} \boldsymbol{\xi} &= (0, -2/3, -4/3, -2) \\ \zeta &= 0.5 \end{aligned} \tag{A21}$$

The true mean structure was specified as in Equation A22. Simulation results based on 100 replications are summarized in Table A1. The bias ranged from -0.058 to 0.033 in the $N = 500$ condition, and it decreased as the sample size increased to $N = 1000$, ranging from -0.026 to 0.032. The RMSE also decreased from (0.187, 0.332) to (0.127, 0.223) as the sample size increased.

$$\begin{pmatrix} 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \\ 10 & 20 & 30 & 40 \end{pmatrix} \tag{A22}$$

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Table A1

Parameter Recovery with Fixed L and $\mathbf{M}_\kappa$.

Parameter	True Value	$N = 500$		$N = 1,000$	
		Bias	RMSE	Bias	RMSE
β_{02}	0.333	0.033	0.206	0.009	0.169
β_{03}	0.667	0.014	0.249	0.013	0.172
β_{04}	1	0.033	0.249	-0.01	0.164
β_{12}	-0.667	0.009	0.226	0.014	0.183
β_{13}	-1.333	-0.018	0.304	-0.011	0.164
β_{14}	-2	-0.041	0.281	0.029	0.223
ξ_{12}	-0.667	-0.056	0.233	-0.026	0.171
ξ_{13}	-1.333	0.008	0.282	-0.013	0.201
ξ_{14}	-2	-0.058	0.332	-0.017	0.210
α_1	-1	-0.006	0.267	-0.013	0.154
α_2	-0.333	-0.002	0.215	-0.008	0.152
α_3	0.333	0.022	0.243	0.002	0.146
α_4	1	-0.024	0.252	0.032	0.199
γ	0	-0.017	0.187	-0.012	0.127
ζ	0.5	-0.005	0.206	0.025	0.128

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

Appendix E. Simulation Under Alternative Data Generating Models

In this section, we evaluate the performance of the LCMA procedure under four alternative data generating models in terms of the variable selection and the parameter estimation accuracy. The alternative models include:

Condition 1 Confounder X and a non-zero γ

Condition 2 Noisy latent class underlying a noisy feature

Condition 3 Unmeasured mediator θ

Condition 4 Mixture Poisson distribution

In each condition, 100 random samples were generated with sample size $N = 500$ and true number of latent classes $L = 4$.

In Condition 1, the effect of a predictor-mediator and mediator-outcome confounder X is included in the data-generating model and is estimated. The true parameter values are specified as Equations A16-A17 and A19-A20. The parameters related to the confounder effect were set as follows.

$$\begin{aligned} \boldsymbol{\xi} &= (0, -2/3, -4/3, -2) \\ \zeta &= 0.5 \end{aligned} \tag{A23}$$

The number of signal indicators was $S = 5$, and the true mean structure was specified as in the $S = 5$ condition in Figure 3. In addition, we included a non-zero $\gamma = 0.2$ value, a non-zero direct effect of the predictor given the latent class mediator and the confounder.

In Condition 2, a noisy latent class variable underlying a noisy feature was generated. This noisy latent class variable was unrelated to both the predictor and the outcome in the generating model. In this condition, we evaluated whether the proposed algorithm correctly selects the signal features despite the presence of a clustering structure underlying a noisy feature. More specifically, the signal latent class variable Ω was

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

generated as a function of the predictor G,

$$P(\Omega_i = \omega \mid G_i = g) = \frac{e^{\beta_{0\omega}+\beta_{1\omega}g}}{\sum_{l=1}^{L} e^{\beta_{0l}+\beta_{1l}g}}. \tag{A24}$$

Then the signal features M_{k^*} were generated as

$$M_{k^*} \mid \Omega = \omega \sim MVN(\mu_\omega, \Sigma_\omega). \tag{A25}$$

Then the final outcome Y was generated as a function of the predictor G and the signal latent class variable Ω, similar to Equation 4. The true parameter values are specified as Equations A16-A17 and A19-A20. The number of signal indicators was $S = 5$ and the true mean structure was specified as in the $S = 5$ condition in Figure 3. The noisy latent class variable Ω^* was generated independently from a Bernoulli distribution.

$$\Omega^* \sim \text{Bernoulli}(0.5). \tag{A26}$$

Then one of the noisy features was generated given the noisy latent class membership as,

$$M_6 \mid \Omega^* \sim MVN(\mu_{\omega^*}, \Sigma_{\omega^*}). \tag{A27}$$

In Condition 3, an unmeasured mediator θ was considered where θ was generated as a function of the predictor G,

$$\theta \mid G \sim N(\mu_g, \sigma_g^2), \tag{A28}$$

where $\mu_g = g, \sigma_g^2 = 0.01$. Then the outcome variable Y was generated as a function of G, Ω, and θ.

$$P(Y_i = 1 \mid G_i = g, \Omega_i = \omega, \theta_i = \theta) = \frac{e^{\gamma g+\alpha_\omega+\theta}}{1 + e^{\gamma g+\alpha_\omega+\theta}}. \tag{A29}$$

The true parameter values are specified as Equations A16-A17 and A19-A20. The number of signal indicators was $S = 5$, and the true mean structure was specified as in the $S = 5$ condition in Figure 3.

In Condition 4, we evaluate the performance of the proposed algorithm under the non-normality assumption. The features were generated under a mixture Poisson

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press

distribution with class-specific rate λ_ω given in Equation A30. Each column represents a latent class. The first five rows represent the signal indicators, and the last five rows are the noisy indicators. All the other true parameter values were set as Equation A16.

$$\mathbf{\Lambda} = \begin{pmatrix} 10 & 2 & 10 & 10 \\ 10 & 10 & 2 & 10 \\ 10 & 10 & 10 & 2 \\ 2 & 10 & 10 & 10 \\ 2 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \end{pmatrix} \tag{A30}$$

The results from the additional simulation with alternative data-generating models are presented in Table A2. The classification accuracy remained reasonably high, with the ARI ranging from 0.79 to 0.91. The false positive rate for selecting noisy indicators ranged from 0.03 to 0.14. The bias in the TIE was small, with relative bias less than 0.1, except in Condition 3 with an unmeasured mediator. When the unmeasured mediator θ was not included in the model, the TIE was overestimated. Similarly, the RMSE of the TIE was highest in Condition 3.

Table A2

Results from the Additional Simulation with Alternative Data Generating Models

Con	ARI	N.Class	N.Ind	FP	TP	TIE		
						Bias	RMSE	95% C.R.
1	0.89	3.34	3.36	0.03	0.64	0.011	0.025	0.92
2	0.79	4.53	3.43	0.14	0.55	0.008	0.024	0.89
3	0.91	3.70	3.15	0.05	0.58	0.026	0.033	0.65
4	0.79	5.15	3.27	0.05	0.61	0.006	0.026	0.88

https://doi.org/10.1017/psy.2025.10038 Published online by Cambridge University Press