JEM JOURNAL of EDUCATIONAL MEASUREMENT NCME national council on measurement in education
DOI: 10.1111/jedm.12412 *Journal of Educational Measurement*

Exploring Latent Constructs through Multimodal Data Analysis

Shiyu Wang and Shushan Wu
University of Georgia
Yinghan Chen
University of Nevada, Reno
Luyang Fang
University of Georgia
Liang Xiao and Feiming Li
Zhejiang Normal University

This study presents a comprehensive analysis of three types of multimodal data-response accuracy, response times, and eye-tracking data-derived from a computer-based spatial rotation test. To tackle the complexity of high-dimensional data analysis challenges, we have developed a methodological framework incorporating various statistical and machine learning methods. The results of our study reveal that hidden state transition probabilities, based on eye-tracking features, may be contingent on skill mastery estimated from the fluency CDM model. The hidden state trajectory offers additional diagnostic insights into spatial rotation problem-solving, surpassing the information provided by the fluency CDM alone. Furthermore, the distribution of participants across different hidden states reflects the intricate nature of visualizing objects in each item, adding a nuanced dimension to the characterization of item features. This complements the information obtained from item parameters in the fluency CDM model, which relies on response accuracy and response time. Our findings have the potential to pave the way for the development of new psychometric and statistical models capable of seamlessly integrating various types of multimodal data. This integrated approach promises more meaningful and interpretable results, with implications for advancing the understanding of cognitive processes involved in spatial rotation tests.

Introduction

Given the prevalence of computer-based assessments and the abundance of digital devices, diverse forms of multimodal data, including response times and eye-tracking data, have become more accessible alongside traditional outcome data like response accuracy. A wide array of psychometric models has emerged, utilizing response accuracy and response time for inference, either independently or in combination. This family of models includes Item Response Theory Models, Cognitive Diagnostic Models (CDM), response time models, and joint models developed within the frameworks of these modeling approaches (e.g., Baker, 2001; Rupp et al., 2010; van der Linden, 2006, 2007). Recent studies in educational and

psychological assessments have increasingly employed eye-tracking technologies to investigate different research problems. These studies utilize eye-tracking data for various purposes, such as uncovering cognitive processes during exams (Kaczorowska et al., 2021; Zhu & Feng, 2015), describing different testing behaviors (Man & Harring, 2023), investigating or gathering valid evidence for item or test construction (Yaneva et al., 2021, 2022), as well as exploring eye movement patterns in relation to different test performances (Hu et al., 2017). In particular, eye-tracking techniques applied to mental rotation tasks have emerged as powerful tools in cognitive research, providing insights into spatial cognition and processing (e.g., Just & Carpenter, 1976; Xue et al., 2017), problem-solving strategies (e.g., Khooshabeh & Hegarty, 2010; Nazareth et al., 2019), and gender differences (Heil & Jansen-Osmann, 2008).

Methodological efforts have also been dedicated to developing novel joint models for these various forms of multimodal data within a structural equation modeling framework (e.g., Zhan et al., 2022). In this framework, different latent constructs are assumed to be reflected by different types of data, focusing on modeling a collective distribution of multiple latent variables. The diagnostic insights provided to students aggregate these latent variables estimated from various data sources concurrently. However, lingering questions from prior research include the nature of relationships between latent constructs identified in different forms of multimodal data, the possibility of redefining diagnostic profiles that encompass students' cognitive processes, attributes, and behaviors by amalgamating individual multimodal data, and the potential to quantify item characteristics using distinct multimodal data.

With these inquiries in mind, this study aims to conduct a multidimensional analysis of three types of multimodal data-response accuracy, response times, and eye-tracking data-collected from a computer-based spatial rotation test. We have chosen these three data types due to their extensive exploration in various applications in educational measurement and assessment, as well as in some existing joint analyses as reviewed above. The specific research questions we aim to address are as follows: (1) How are the latent constructs manifested in eye-tracking features related to participants' fine-grained spatial rotation skills, as estimated from a specific CDM, the fluency CDM (Wang & Chen, 2020), based on response accuracy and response time? (2) How can eye-tracking features reflect item characteristics in addition to those identified in the fluency CDM?

A methodological challenge in this multidimensional exploratory analysis pertains to managing the high-dimensional eye-tracking variables. Unlike many prior studies that concentrate solely on a limited set of metrics gleaned from eye-tracking studies (e.g., Zhan et al., 2022; Man & Harring, 2023), our study utilizes a three-step feature selection strategy, accompanied by a comprehensive data preprocessing guide. This strategy enables us to effectively identify the most predictive and meaningful eye-tracking features from a pool of 100 variables generated by the Tobii Pro Lab software. Moreover, we introduce a comprehensive methodological framework that leverages an array of statistical and machine learning techniques to unveil the latent structure inherent within these selected variables. Consequently, from a methodological standpoint, this study contributes to the management of diverse types of high-dimensional multimodal data. From a practical and theoretical

perspective, the answers to our research questions could potentially pave the way for the development of novel psychometric and statistical models capable of handling various types of multimodal data in a more integrated manner, ultimately yielding more meaningful and interpretable results.

Experiment and Data

The experiment utilized the spatial rotation learning program developed by Wang et al. (2020) to assess four mental rotation skills: 90 degree and 180 degree of rotation along the x axis and y axis. These four skills are referred as the four attributes in previous studies, which utilize this data set to conduct CDM analysis (Wang et al., 2020; Zhang & Wang, 2018). This learning program consists of two testing modules and two learning modules, with the learning modules positioned between the two testing modules. The psychometric properties of this learning program can be found in Wang et al. (2020).

The experiment was conducted in the eye-tracking laboratory of a University in China. A computer equipped with a Tobii Pro Spectrum eye tracker was used to collect data from participants. This eye tracker does not require wearing and allows for minor head movements by the participants. The spatial rotation learning program was presented on a display with a resolution of 1,920-1,080 and a screen refresh rate of 60 Hz. Each question is displayed on a separate webpage, and participants are not permitted to revisit or revise their responses to a question after they have submitted it. The participants were positioned at a distance of roughly 60 cm from the screen.

Experiment Procedures

Ninety-one undergraduate students were randomly selected to participate in the experiment, all possessing normal or corrected-to-normal vision without any color blindness or color vision impairments. Prior to commencing the experiment, participants were briefed on the instructions and assured a predetermined reward upon completion. At the beginning of the experiment, the experimenter provided a clear explanation of its purpose and pertinent instructions. Participants were explicitly instructed not to use their hands or other objects to obstruct the space between the eye tracker and their eyes. Furthermore, they were guided through the process of signing an informed consent form before engaging in the experiment. Subsequently, the eye tracker was adjusted to ensure correct seating, maintaining a consistent distance of 60-65 cm between the device and participants' eyes. According to the Tobii Pro Spectrum eye tracker specifications, the acceptable angle error was kept below 1.0°. Throughout the experiment, participants navigated spatial rotation test questions displayed on the screen at their own pace. To avoid rushed responses and forced guessing, no time limit was imposed. On average, the experiment duration was approximately 40 minutes.

Variables and Data

This study exclusively utilizes data from the initial module of the spatial rotation learning program which consists of 10 questions for subsequent analysis. The rationale behind this choice is that the study's primary objective is to investigate the

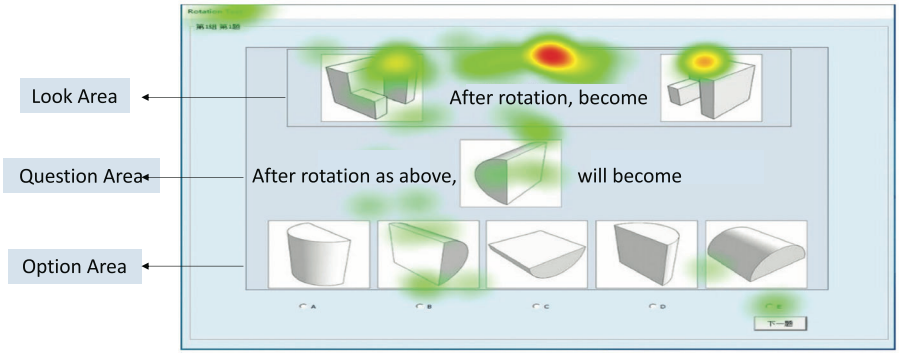

Figure 1. Three specific AOIs from an example question shown in the question panel.

cognitive processes and attributes that signify an individual's present capability or performance within an assessment, rather than assessing their learning process.

Eye-tracking data throughout the experiment were captured using Tobii Pro Studio software, and response accuracy and response time were recorded in text files from the spatial rotation learning program. For each participant, the response accuracy is defined as the binary response vector, indicating whether their response to each question is right or wrong, to all questions. The response time to a question is defined as the time they spend to complete that question. There are 100 eye-tracking variables that were collected in total. The following subsection provides details on these eye-tracking variables.

Eye-tracking variables. The eye-tracking variables are defined based on four major metrics on eye movement. They are fixations, saccades, glances, and visits. Fixations are defined as the periods of time where the eyes are relatively still and a sequence of raw gaze points where the estimated velocity is relatively low in "Tobii Pro Lab." Saccades, on the other hand, are responsible for moving one's eyes to different objects or areas of interest within their field of vision, which is the opposite of fixation. The other two measures, visits and glances, are associated with general looking behavior and attention to Areas of Interest (AOIs). An AOI refers to a specific region or area within a visual stimulus or scene that researchers define and track to analyze participants' eye gaze behavior. The visit associated with an AOI corresponds to all the data between the start of the first fixation inside the AOI to the end of the last fixation in the same AOI. The glance associated with an AOI is all data (even saccades, blinks, or invalid gaze data) from the first saccade leading into the AOI until the last fixation inside the AOI.

In our study, in addition to a single large AOI defined based on the overall region of the interface, we subjectively define three smaller AOIs in one screen, the look area, the question area, and the option area, as shown in Figure 1. The rationale of defining these three specific AOIs is as follows. First, eye movements in the look area are important because they provide insight into how a participant is studying the mental rotation shown in the example object. Second, the question area contains the instructions for the task and the presentation of a new object. Participants need to

Table 1
A Summary of Eye-Tracking Features

Metrics	Statistics	AOI
	Total duration (ms)	Look
Glance	Average duration (ms)	Question
	Maximum duration (ms)	Option
Fixation	Minimum duration (ms)	Overall region
	Number (count)	
Visit	Time to first (ms)	
	Duration of first (ms)	
Saccade	Number (count)	Look
	Peak Velocity (degree/ms)	Question
	Amplitude (degree)	Option
	Direction (degree)	Overall region

encode the stimulus and identify the orientation of the stimulus in order to correctly perform the same mental rotation task as shown in the example in the look area. Finally, the information in the option area is critical. Participants need to figure out the ending position of the stimulus in the question area following the same rotation that they figure out from the look area. Based on a specific AOI, different variables are defined for the four eye movement metrics, which are documented in Table 1.

Data preprocessing. During experiments, eye tracking data can be affected by factors like subject blinking, unstable head positioning, or hand movements obstructing the eye tracker. The eye tracker's sampling rate percentage is determined as the ratio of the number of accurately identified eye tracking samples to the theoretical maximum value, typically resulting in a 5%-10% data loss (Lab, 2023). This loss rate correlates with the experiment's duration; the longer the study, the more challenging it becomes for subjects to maintain a stable head posture. Given the 30- to 40-minute duration of our study, some data loss due to subjects' fatigue or head instability is inevitable. Therefore, in balancing the need for an adequate subject pool and the quantity of valid eye tracking data per subject, we pragmatically selected subjects with a sampling rate exceeding 70% as our final valid participants. As a result, data from 21 participants were excluded, leaving 70 participants for data analysis.

We initially excluded a set of variables from the following analysis through a specific procedure. First, we removed variables related to "Time to first" as they only capture the date-time value of initial eye movement metrics, which does not effectively reflect the promptness of subject responses to items. Second, our preliminary quantitative analysis of two sets of eye-tracking metrics, namely "Visits" and "Glances," revealed a significant correlation between them. Among the 24 metrics documenting "Visit" behavior, 22 variables exhibited a strong correlation (Pearson's correlation above.90) with their corresponding metrics in "Glances" behavior. Consequently, we opted not to analyze variables related to "Visit" metrics. Finally, we excluded variables with a high incidence of missing values, defined as those exceeding 5% of the total number of observations. Appendix 1, Tables A1-A3 display the

number of variables removed after each step in the data preprocessing procedure. The full list of variables after the data preprocessing step can be found in Appendix 1, Table A4. In summary, 65 eye-tracking variables remain after completing the data preprocessing step.

Methods

Eye Tracking Feature Selection

In order to select the most predictable and meaningful eye-tracking features for response accuracy and response times for subsequent statistical analysis, we propose a three-step feature selection strategy utilizing the generalized linear mixed-effects model.

To start with, we denote Y_{ij} as the response accuracy variable when ith participant answering jth question, taking values from $Y_{ij} = \{0, 1\}$, where 0 means that the answer is wrong and 1 otherwise. A logistic link function model is used to connect the expectation of the binary response given the data with a linear combination of the predictors as follows:

$$P\big(Y_{ij} = 1 \mid \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{b}\big) = h(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}\boldsymbol{b}),$$
$$Y_{ij} \sim Ber(P\big(Y_{ij} = 1 \mid \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{b}\big)),$$

where $h(\cdot)$ is a link function, in this case, $h(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}\boldsymbol{b}) = \frac{\exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\boldsymbol{Z}_{ij}\boldsymbol{b}\}}{1+\exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\boldsymbol{Z}_{ij}\boldsymbol{b}\}}$. $\boldsymbol{X}_{ij}$ is the eye tracking features vector of ith participant answering jth question in the eye tracking feature matrix, $\boldsymbol{\beta}$ is an unknown vector of fixed effects, $\boldsymbol{Z}_{ij}$ is the design matrix documenting which question the participant is answering, and $\boldsymbol{b}$ is an unknown vector of random effects. In the mixed effects model, the random effects $\boldsymbol{b}$ are assumed to be multivariate normal with zero mean and covariance $\boldsymbol{\Sigma}$. The model parameters are estimated by maximizing the likelihood function.

$$\max_{\boldsymbol{\beta}\in\mathbb{R}^L, \boldsymbol{b}\in\mathbb{R}^J} \left[\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \big\{ Y_{ij} \log P\big(Y_{ij} = 1 \mid \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{b}\big) \right.$$
$$\left. + \big(1 - Y_{ij}\big) \log \big(1 - P\big(Y_{ij} = 1 \mid \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{b}\big)\big) \big\} \right].$$

In literature, response time is commonly modeled as the log normal distribution (van der Linden, 2006; Wang et al., 2020). We conduct the logarithm transformation for the response time variable $\mathcal{T}_{ij}$, then fit $\log \mathcal{T}_{ij}$ into a linear mixed effects regression model, where the conditional probability of log response time given the predictors $\boldsymbol{X}_{ij}$ is:

$$\Pr(\log\big(\mathcal{T}_{ij}\big) \mid \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{\big[\log\big(\mathcal{T}_{ij}\big) - (\boldsymbol{X}_{ij}\tilde{\boldsymbol{\beta}} + \boldsymbol{Z}_{ij}\tilde{\boldsymbol{b}})\big]^2}{2\sigma^2} \right\}.$$

The model parameters are estimated by maximizing the likelihood function:

$$\max_{\tilde{\boldsymbol{\beta}}\in\mathbb{R}^L,\tilde{\boldsymbol{u}}\in\mathbb{R}^J}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\left[\log\left(\mathcal{T}_{ij}\right)-\left(\boldsymbol{X}_{ij}\tilde{\boldsymbol{\beta}}+\boldsymbol{Z}_i\tilde{\boldsymbol{b}}\right)\right]^2\right].$$

The parameters $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{b}}$ are estimated using the lme4 R package (Bates et al., 2014).

Our proposed three-step feature selection strategy unfolds as follows. Initially, we address the multicollinearity issue, given the considerable number of eye-tracking features (65). We employ all 65 features as independent variables and construct an initial mixed-effects model for both response accuracy and response times. Given that eye-tracking features vary in scale, we standardize each variable prior to incorporating them into the generalized linear mixed-effects model. Through the initial mixed effect model results, we are able to calculate the Variance Inflation Factor (VIF) and p-values for each feature. VIF quantifies the correlation of a variable with a linear combination of other predictors and is computed using the formula $\text{VIF}_x = \frac{1}{1-R_x^2}$, where R_x^2 is the R^2 value from regressing the predictor x on all other predictors by mixed effects models. Typically, VIF values exceeding 10 indicate significant multicollinearity concerns, prompting remedial action such as removing the affected variables (Weisberg, 2005). In our current analysis, we adopt a more conservative threshold of VIF $>$ 200. Variables with high VIF values (VIF $>$ 200) and p-values surpassing.5 for response accuracy and.1 for response time in the mixed-effects model are iteratively eliminated. Following this initial step, we effectively reduce the feature set to a manageable size. Subsequently, we refit the mixed-effects models for both response times and accuracy, utilizing these selected eye-tracking features. Further refinement involves selecting features significant at a.05 significance level for response accuracy and.01 for response times. The final step involves scrutinizing the correlation among the features selected in Step 2. Features exhibiting moderate pairwise correlations with others (larger than the third quantile of the absolute values of pairwise correlations) are pruned from the selection to ensure a robust feature set.

The First-Order Hidden Markov Model

To uncover the underlying cognitive processes involved in visual perception and attention, we use a hidden Markov model (HMM) to analyze the selected eye-tracking features from "Eye-Tracking Feature Selection" section. The motivation for using HMMs in eye-tracking research lies in the fact that eye movements are inherently sequential in nature and often involve a dynamic interplay between bottom-up sensory input and top-down cognitive influences. HMMs are particularly suited for modeling sequential data because they can capture temporal dependencies in the data and allow for the modeling of hidden (i.e., unobserved) states that may correspond to different cognitive processes. In fact, not only for eye-tracking data (Xue et al., 2017), HMMs have been utilized to discover hidden states for a variety of problem-solving process data (Wang et al., 2023; Xiao et al., 2021). The particular HMM we considered is the first-order HMM. This means that the model's current state depends

only on its immediate predecessor, which simplifies the state transition dynamics but is appropriate given the context of our analysis.

To start with, we denote the selected eye-tracking features by $(\mathbf{O}_{1:J} = (O_1^1, \ldots, O_1^M, O_2^1, \ldots, O_2^M, \ldots, O_j^1, \ldots, O_j^M, \ldots, O_J^1, \ldots, O_J^M)$ for M-variate time series of length J, and $\mathbf{O}_j$ is short for $(O_j^1, \ldots, O_j^M)$. In our case, J is the number of questions in the test and M is the number of selected eye-tracking features from the eye-tracking feature selection process. In the dependent mixture model (Visser &Speekenbrink, 2010), each observation is distributed as a mixture of the p states. The time dependence between the observations can be modeled by the transition probability between hidden states. Then the joint likelihood function of the observations $\mathbf{O}_{1:J}$ given the hidden states $\mathbf{S}_{1:J} = (S_1, \ldots, S_J)$ and the model parameters $\boldsymbol{\theta}$ can be written as:

$$P(\mathbf{O}_{1:J}, \mathbf{S}_{1:J} \mid \boldsymbol{\theta}) = \pi_{\boldsymbol{\theta}_1}(S_1) \prod_{j=1}^{J-1} P_{\boldsymbol{\theta}_2}\big(S_{j+1} \mid S_j\big) P_{\boldsymbol{\theta}_3}\big(\mathbf{O}_j \mid S_j\big), \tag{1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3})$ is the general parameter vector consisting of three subvectors with parameters for the prior model, transition model, and response models, respectively. An appropriate density function needs to be determined based on the nature of the dependent variables. The model is estimated by maximizing the joint likelihood function using the expectation-maximization (EM) algorithm in the statistical package depmixS4 (Visser & Speekenbrink, 2010). The number of hidden states is an important hyperparameter in the HMM, and the estimation of the remaining model parameters relies on a fixed number of hidden states. To better interpret the model results, we need to select an appropriate model. One important criterion that guides us to select hyperparameters is the Bayesian information criterion (BIC), $\text{BIC} = r\ln(N) - 2\ln(\hat{\mathcal{L}})$, where r is the number of parameters to be estimated, N is the number of observations, and $\hat{\mathcal{L}}$ is the maximized value of the likelihood function. The model with a lower BIC value is preferred.

The Fluency Cognitive Diagnostic Model

In addition to the hidden states discovered from the HMM, we also estimate the latent attribute profiles, using the response accuracy and response times, through a recently proposed fluency cognitive diagnostic model (CDM) (Wang & Chen, 2020). The fluency CDM provides fine-graded diagnostic information regarding a person's mastery of assessed skills in an assessment by jointly analyzing the response times and response accuracy. Specifically, assuming an assessment consists of J items measures K attributes. The item-attribute association is documented through a Q matrix with binary component q_{jk}, indicating whether item j measures attribute k ($q_{jk} = 1$) or not ($q_{jk} = 0$). A sample of N participants complete the assessment, and the response accuracy and response time from participant i to item j are denoted as Y_{ij} and L_{ij}. The fluency CDM model defines a latent attribute with three levels: nonmastery, partial-mastery, and mastery (i.e., fluency). Thus, if we denote $\boldsymbol{\alpha}_i$ as the latent attribute profile of the participant i, then $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{ik})^T$. In this case, $\alpha_{ik} = 0$ indicates the nonmastery level, reflecting participant i has low response accuracy; $\alpha_{ik} = 1$ denotes the partial-mastery level, meaning participant i has high accuracy

but low speed on correct answers; $\alpha_{ik} = 2$ denotes the mastery level, which indicates students not only have high accuracy on responses but also fast speed when answering questions correctly, thus also presented as the fluency level. Within this framework, the fluency CDM employs two measurement models, one for response time and another for response accuracy, to assess both components.

The measurement model for response accuracy is described by Equation 2, which classifies the participants into three types of correct response probabilities. Specifically, this model first defines the ideal response of a participant i to item j as $\eta_{ij} := \eta_{ij}(\boldsymbol{\alpha}_i, \boldsymbol{q}_j) = 1_{\{\forall k, q_{jk}=1, \alpha_{ik}=q_{jk}\}} + 1_{\{\forall k, q_{jk}=1, \alpha_{ik}>q_{jk}\}}$. When $\eta_{ij} = 0$, it denotes a participant lacks mastery in any of the required attributes for that item, so one may have g_j probability to guess this item correctly. When $\eta_{ij} = 1$, it indicates that a participant has partial mastery of all required attributes, but at least one attribute has not reached the fluency level. In this case, one can have $1 - s_{1j}$ probability to answer this item correctly. Finally, when $\eta_{ij} = 2$, it suggests that a participant has mastered all the required attributes, all these attributes have reached the mastery level. Thus, in this case, one can correctly answer this item with probability $1 - s_{2j}$. A monotonicity assumption is imposed such that $0 < g_j < 1 - s_{1j} < 1 - s_{2j} < 1$.

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_i) = \begin{cases} g_j, & \text{if } \eta_{ij} = 0 \\ 1 - s_{1j}, & \text{if } \eta_{ij} = 1 \\ 1 - s_{2j}, & \text{if } \eta_{ij} = 2. \end{cases} \tag{2}$$

The measurement model for response time follows a log-normal distribution (van der Linden, 2007), depending on whether $Y_{ij} = 1$ or $Y_{ij} = 0$. Essentially, when $Y_{ij} = 1$, the model leverages response time data to differentiate between a student with partial mastery and one with fluency in correctly answering an item. Wang and Chen (2020) introduced two response time models, each based on whether it assumes that students have consistent or varying speeds when completing items correctly or incorrectly. Here we present the response time model assuming that students have the same base speed in the following Equation 3

$$\log(L_{ij}) \sim \begin{cases} N\left(\gamma_j - (\tau_i + \phi_i \times g(\boldsymbol{\alpha}_i, \boldsymbol{q}_j)), \frac{1}{a_j}\right) & \text{if } Y_{ij} = 1 \\ N\left(\gamma_j - \tau_i, \frac{1}{a_j}\right) & \text{if } Y_{ij} = 0 \end{cases}, \quad \tau_i \sim N\left(\mu_\tau, \sigma_\tau^2\right), \tag{3}$$

In this model, τ_i is the base speed of participant i, indicating participant i's initial speed of answering questions. γ_j is the time intensity parameter, it measures the overall times when solving item j. a_j^2 is the time discrimination parameter, which captures the variance of log-response times given τ_i and γ_j, indicating the sensitivity of item j to differentiate students- base speeds. The covariate $g(\boldsymbol{\alpha}_i, \boldsymbol{q}_j)$ is a monotonic function measuring how latent attribute profile $\boldsymbol{\alpha}_i$, influences response times on answering item j correctly. ϕ_i is student $i's$ speed change parameter, controls the effect of function $g(\boldsymbol{\alpha}_i, \boldsymbol{q}_j)$, and is constrained to be positive. Wang and Chen (2020) provide three different forms of $g(\boldsymbol{\alpha}_i, \boldsymbol{q}_j)$, suggesting the specific forms need to be constructed based on a specific assessment structure. The model estimation

procedures and identifiability issues have been addressed by Wang and Chen (2020). For this study, we followed their developed MH-Gibbs algorithm to estimate the fluency CDM.

Analysis for Two Research Questions

Building upon the findings from "The First-Order Hidden Markov Model" and "The Fluency Cognitive Diagnostic Model" sections, we investigate the relationship between the hidden states estimated by the first-order HMM model and the attribute profile derived from the fluency CDM from two perspectives.

Initially, we investigate whether the empirical transition matrices of hidden states vary across different levels of skill mastery between adjacent items. This analysis sheds light on whether the transition matrix of the HMM depends on varying skill mastery levels.

Subsequently, we scrutinize how state trajectories diverge among participants exhibiting differing levels of skill mastery. A hidden state trajectory of participant is defined as their hidden states across 10 items, as in our case, one item represents one time point. To accomplish this, we categorize participants into three distinct groups based on their latent attribute profiles, representing low, medium, and high skill levels. We then examine the distribution of hidden states across the 10 items for each of these latent profile groups. We also showcase the variation in hidden state trajectories among participants sharing the same latent attribute profile identified through fluency CDM.

Lastly, to gain insights into item characteristics, we analyze the estimated item parameters from the fluency CDM model and assess the distribution of hidden states for each item. Additionally, we compute the correlation between the proportions of hidden states for each item and the fluency CDM item parameters. To facilitate understanding, we employ a multivariate plotting approach to visualize each item characterized by fluency CDM item parameters alongside hidden state proportions.

Results

Eye-Tracking Feature Selection Results

Through the three-step feature selection strategy, a total of 10 eye-tracking features remained, and these are detailed in Table 2. The eye-tracking variables removed in the initial step and those are selected in the second step are documented in Appendix 2, Tables A1-A4.

Table 2 includes the associated regression coefficients and various statistics for each of the eye-tracking variables. The chosen variables, based on both response accuracy and response time, pertain to distinct types of eye movements, namely fixation, glances, and saccades. Furthermore, these selected variables encompass all regions of interest, including look, question, option areas, and the overall interface region. These selected eye-tracking variables are consistent with the major fixation and saccade-related features, discussed by recent eye-tracking studies (Hu et al., 2017; Kaczorowska et al., 2021; Xue et al., 2017; Yaneva et al., 2021; Zhu & Feng, 2015).

Three response accuracy-related variables have been selected: the average duration of Glances in the option area, the maximum duration of Glances in the question area,

Table 2
The Selected Eye-Tracking Features

Response	Variables (Unit)	Coef	Min	Mean	Max
Accuracy	Average duration of Glances option (ms)	−.5499	104	1,130	9,074
	Maximum duration of Glances question (ms)	−.4833	54	2,681	23,264
	Maximum peak velocity of saccades (degree/ms)	0.4998	30.3	370.5	1061.2
RT	Number of fixations look (count)	.7387	0	49.43	1,124
	Maximum duration of fixations look (ms)	.0670	92	511.5	2,041
	Average duration of fixations option (ms)	−.0529	70	193	421
	Maximum duration of fixations question (ms)	.1226	38	517	6,041
	Minimum duration of fixations question (ms)	−.0334	33	70.79	200
	Number of Glances	.2098	1	2.942	35
	Minimum duration of Glances (ms)	−.3235	33	18326.5	150452.0

and the maximum peak velocity of saccades from the entire screen. Analysis of the estimated regression coefficients suggests that an individual with a longer average duration of glances in the option area and an extended maximum glance duration in the question area is likely to exhibit lower response accuracy. Conversely, a person with higher maximum peak velocities may demonstrate higher response accuracy, even after accounting for other variables. A substantial average duration of glances in the option area may signify heightened participant attention, especially in discerning the orientation of objects within that region. Meanwhile, a longer maximum glance duration may indicate that an individual requires more time for comprehension. The presence of high peak velocities could suggest rapid gaze movements, potentially reflecting confidence in the chosen answer and a desire to progress.

Seven variables are selected based on response time. A higher number of fixations in the look area may indicate that participants require additional time for content comprehension. Extended fixation durations may suggest heightened cognitive processing or increased concentration on the content. Specifically, prolonged fixations in the option area indicate participants' meticulous attention and consideration for each option. Shorter minimum fixation durations may hint at instances where participants overlook certain parts of the question, potentially leading to faster, albeit potentially inaccurate, responses. An increased number of glances suggests that participants frequently check and reevaluate their answers or the content, contributing to prolonged response times. Shorter glance durations may indicate swift evaluations or dismissals of specific options or content.

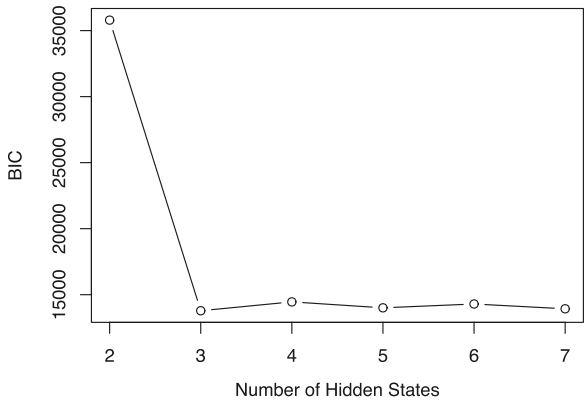

Figure 2. The BIC values of HMM with different states.

Table 3
Transition Probability of Three Hidden States from the HMM

	to S1	to S2	to S3
from S1	.500	.489	.012
from S2	.252	.629	.119
from S3	.132	.605	.264

Table 4
Initial Probabilities of Three Hidden States from the HMM

State 1	State 2	State 3
.032	.300	.668

In summary, these chosen variables provide valuable insights into participants' visual and cognitive processes within specific and overall areas of interest.

HMM Results

Because the selected eye-tracking features have different scales, as shown in Table 2, for this study, we applied a box-cox transformation to each selected eye-tracking feature. After the transformation, the eye-tracking features are significantly normal and validated by the normality tests. Thus, the normal density is selected for the HMM. As illustrated in Figure 2, the HMM with three hidden states exhibits the lowest BIC value. In addition, beyond the BIC value, we take into account the model's interpretability. Considering both factors, we ultimately choose the HMM with three hidden states. The transition probabilities for these states are presented in Table 3. The initial probabilities are presented in Table 4. The emission probability of the multivariate response variable given each hidden state follows multivariate Gaussian distribution, and the mean and standard deviation (*SD*) of the emission probability is given in Table 5. Upon analysis of the table, we can conclude that

Table 5
Emission Probabilities of the Multivariate Response Variables Given Each of Three Hidden States

Mean (*SD*)	State 1	State 2	State 3
Variable 1	10.064 (.800)	11.521 (1.394)	435.393 (106.358)
Variable 2	9.982 (.945)	11.674 (1.844)	364.078 (145.409)
Variable 3	9.767 (1.375)	9.748 (1.888)	265.334 (157.868)
Variable 4	54.706 (42.639)	12.668 (1.452)	27.308 (3.066)
Variable 5	56.619 (68.806)	12.127 (1.882)	24.933 (3.716)
Variable 6	17.807 (14.458)	10.456 (1.507)	24.536 (4.707)
Variable 7	8.841 (.807)	3.489 (.213)	.479 (.541)
Variable 8	8.453 (1.117)	3.433 (.000)	.533 (.573)
Variable 9	7.361 (.923)	3.425 (.266)	.344 (.428)
Variable 10	47.996 (26.379)	47.528 (28.003)	44.864 (22.967)

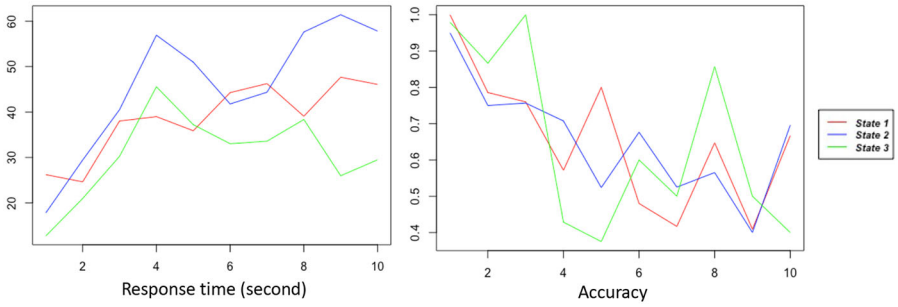

Figure 3. The mean response time and mean response accuracy of three hidden states on each question.

when a participant is in State 1 (S1), there is a tendency to either remain in this state or transition to State 2 (S2). If a participant is in State 2, there is a preference to stay in State 2. For participants in State 3 (S3), there is a tendency to transition to State 2 or remain in State 3 compared with transitioning to State 1.

To better interpret the three hidden states, we plotted the mean response time, mean response accuracy, and the selected eye-tracking features across different items. Figure 3 presents the documented mean response time and response accuracy for three hidden states. These values are computed based on the data of participants who belong to the same state. When considering individual items, participants in State 2 tend to exhibit longer response times compared to those in States 1 and 3. However, their response accuracy is generally similar to that of participants in State 1. On the other hand, participants in State 3 typically demonstrate shorter overall response times and higher response accuracy for items assessing a single skill. For items measuring multiple skills, their response accuracy is lower, except for items 7 and 8, where they stand out as exceptions when compared to the other two states.

The mean values of the eye-tracking statistics gathered from the three states for each item across four Areas of Interest (AOIs) are presented in Figures 4–7.

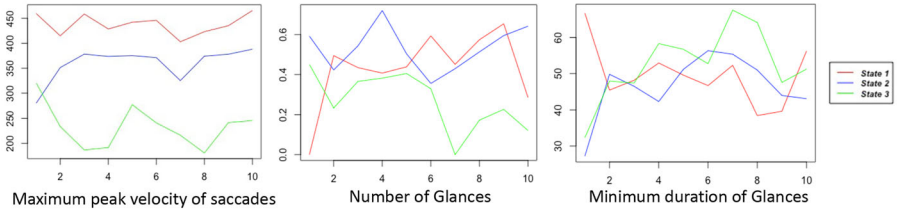

Figure 4. The mean of eye tracking feature statistics on the whole screen.

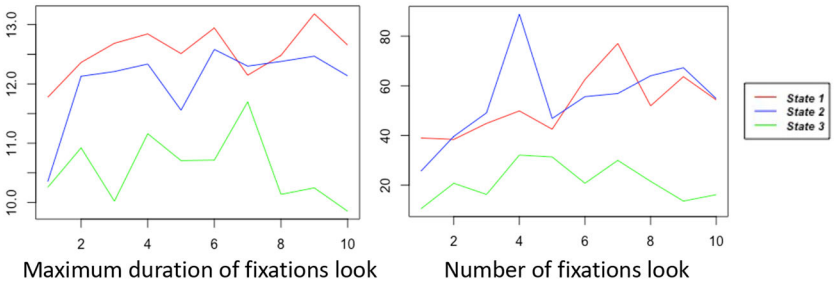

Figure 5. The mean of eye tracking feature statistics in the AOI look.

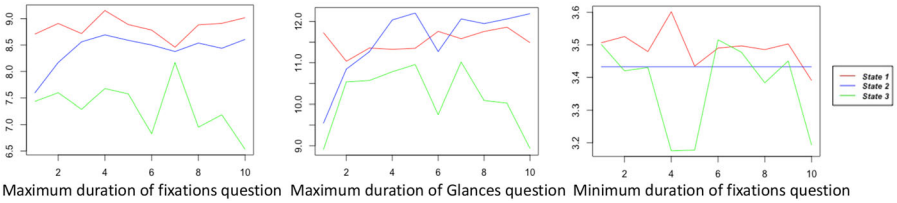

Figure 6. The mean of eye tracking feature statistics in the AOI question.

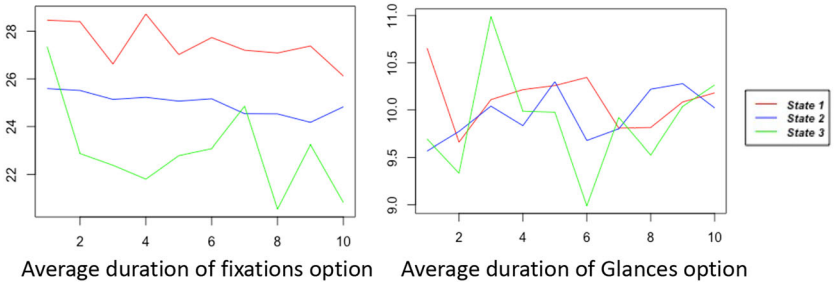

Figure 7. The mean of eye tracking feature statistics in the AOI option.

When examining the 10 eye-tracking features within these AOIs, a consistent trend emerges: participants in State 3 consistently exhibit the lowest values for most eye-tracking features. State 1 and State 2, on the other hand, display a more similar performance compared to State 3. In particular, State 1 generally exhibits the highest values for many of the eye-tracking features. Based on the meaning of the selected

eye-tracking features ("Eye-Tracking Feature Selection Results" section), we interpret the three hidden states based on the eye movement characteristics as follows:

- **State 1**: Participants in State 1 exhibit the highest maximum peak velocity saccades across the entire screen, the longest maximum fixation duration in the look and question area, relatively large number of fixation in the look area, and the lengthiest average fixation duration in the option area across all 10 items. These statistics imply that these participants may possess a keen attention to detail, excellent concentration, and the ability to make rapid decisions when solving problems. However, it is worth noting that these data may also suggest that these participants could face challenges in the visualization process. We define this State as "Precision Navigation."
- **State 3**: Participants in State 3 display characteristics that are notably distinct from those in State 1. They exhibit the smallest maximum peak velocity of saccades across the entire screen, the shortest maximum fixation duration in the look and option areas, the fewest fixations in the look area, and the shortest maximum duration of glances in the question area. These statistics imply that these participants excel in rapid information processing, maintain focused attention on critical details, and exhibit a preference for fast-paced environments. Based on these characteristics, we define State 3 as "Quick Analyzing."
- **State 2**: Participants in State 2 in general have similar eye-tracking features as State 1, they are more like the blend of the other two states, balancing between searching and processing visual information. We define this state as "Balanced Precision Navigation."

Latent Attribute Profiles from Fluency CDM Model

The fluency CDM described in "The Fluency Cognitive Diagnostic Model" section was used to fit the response accuracy (binary response data) and response times from 70 participants to the 10 questions. Given the relatively small sample size, the covariate $g(\boldsymbol{\alpha}_i, g)$ was chosen with a simple form of $1_{\{\eta_{ij}=2\}}$., which differentiates the speed at the highest attribute level from the other two. The MH-Gibbs algorithm converged after 4,500 iterations based on the Gelman-Ruin proportional scale reduction factor (PSRF; Gelman & Rubin, 1992). We thus used a chain length of 20,000 and the first 5,000 iterations as burn-in. The trace plot and auto-correlation plot of the key model parameters were also produced to monitor model convergence. We also conducted Bayesian posterior productive check to evaluate model-data fitting (Gelman et al., 1995). We define the testing statistics used in the posterior predictive checking as the sum of response scores and sum of response times. The posterior predictive p value (PPP) is calculated as $\frac{\sum_{i=1}^{N}\sum_{r=1}^{R}(\sum_{j=1}^{J} Y_{ij}^{posterior(r)} \geq \sum_{j=1}^{N} Y_{ij})}{NR}$, where R represents the MCMC chain length, N is the sample size, and J is the number of test questions. $Y_{ij}^{posterior(r)}$ denotes the predicted posterior data for the ith person's response accuracy or response time to the jth question in the rth MCMC iteration generated by the fluency CDM. Y_{ij} represents the corresponding observed

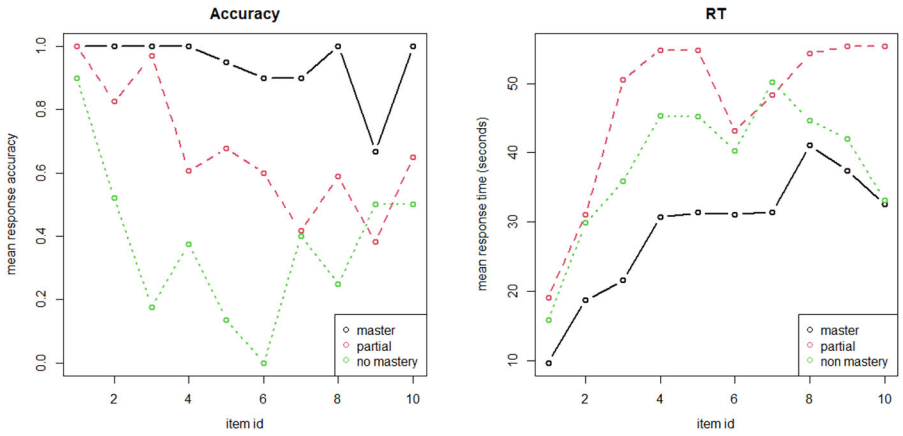

Figure 8. The mean response accuracy and response times for three latent groups based on fluency model.

data. Following Gelman et al. (1995), a PPP value close to.5 indicates minimal disparity between observed and predicted values. A PPP value exceeding.95 or falling below.05 suggests poor model-data fitting. In this case, the fluency CDM model exhibits a PPP value of.491 for response accuracy and.601 for response times, indicating reasonable model-data fitting.

A total of 80 latent attribute profiles were estimated using fluency CDM. To better interpret the meaning of different latent profiles, we check the participants' classification results on each item. More specifically, based on the estimated latent profile, the fluency CDM model classifies a participant into one of three groups: nonmaster, partial master, or master. We then calculate the mean response accuracy and mean response times for each of the three latent groups on each item. The results are documented in Figure 8. It is very clear that the master group has the highest response accuracy in responding to each question and they also have the fastest speed in responding to most items. The mean response accuracy for the partial master group is the next, and this group also tends to have the slowest speed in responding to each question. The nonmaster group answers each with the lowest accuracy and fastest speed.

The Relationship between Fluency CDM Latent Attribute Profiles and Hidden States from Eye-Tracking Features

Hidden states transition matrix and skill mastery. We first report the results about whether the empirical transition matrices of hidden states vary among different types of skill mastery levels of two adjacent items. We created empirical transition matrices for the following three cases: (1) when the skill mastery level (η_{ij}) remains consistent across two adjacent items; (2) when η_{ij} shifts from a high to low value; and (3) when η_{ij} transitions from low to high. A total of nine empirical transition matrices are documented in Figure 9.

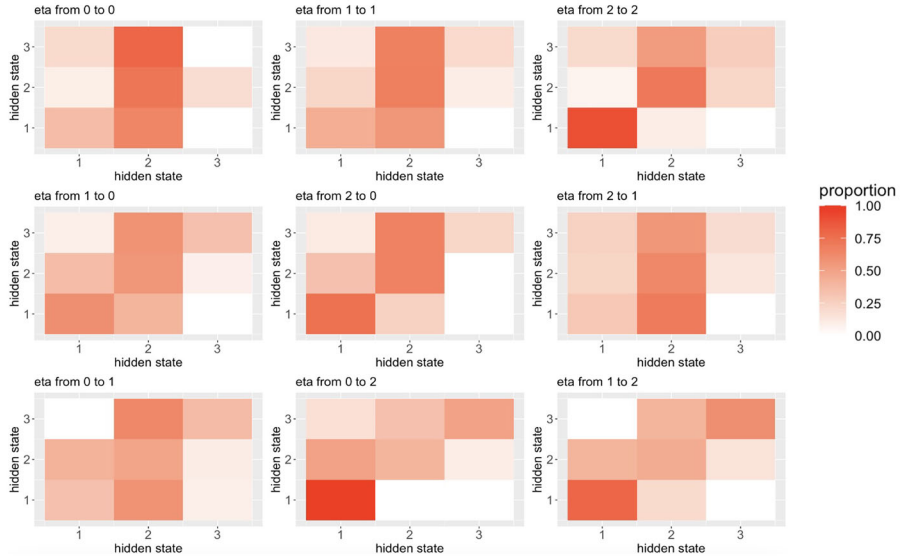

Figure 9. The conditional transition matrices.

Among these 9 empirical transition matrices, most of them align with the estimated transition matrix in Table 3. However, we do observe several differences. First, when examining individuals whose skill levels remain unchanged between two consecutive items, those with ideal responses of (0 to 0) or (1 to 1), and who are in State 1 during the current item, show a notable tendency to transition to State 2 for the next item. If these individuals are in State 3 during the current item, they display a strong inclination to transition to State 2 for the subsequent item. However, for participants whose ideal responses shift from 2 to 2, those in State 1 during the current item typically remain in State 1, while those in State 3 have a slightly higher chance of transitioning to State 2 compared with remaining in State 3 or transitioning to State 1.

Second, individuals whose ideal responses shift from a high value to a low value between two consecutive items, as in (1 to 0) and (2 to 0), exhibit transition patterns closely resembling the overall transition matrix found in Table 2. However, for those whose ideal responses shift from 2 to 1, if they are in State 1 during the current item, they tend to transition to State 2 for the next item.

Third, for participants whose ideal responses change from a low value to a high value between two adjacent items, intriguing patterns emerge. For those shifting from 0 to 2, which constitutes a substantial increase, they remain largely in State 1 if they are in State 1 during the current item. This contrasts with the overall pattern, which features a.489 transition probability from State 1 to State 2.

A final note is for participants whose ideal response change from 0 to 2, from 1 to 2, or remain in 2 between consecutive items, all have the highest rate of staying in State 1.

Hidden states distribution and skill mastery. To present the distribution of hidden states across 10 items for different latent attribute profile groups, we categorize

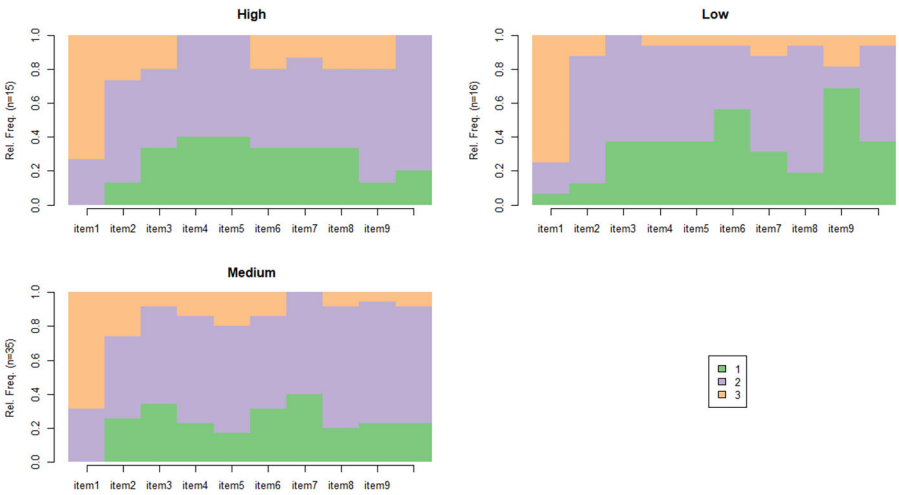

Figure 10. The distribution of hidden states of three latent profile group.

participants into three distinct groups based on their latent attribute profiles, representing low, medium, and high levels. The low latent attribute profile group consists of 16 participants, characterized by ideal response scores of 0 or 1 for most items. The medium latent attribute profile group comprises 35 participants, where a majority of the participants exhibit ideal response scores of 1 across their items. In contrast, the high latent attribute profile group includes 15 participants, with most individuals in this group displaying ideal response scores of 2 for the majority of items.[1]

The distribution of hidden states across 10 items for the three latent profile groups is illustrated in Figure 10. A common feature among these groups is the predominant presence of participants in State 2 (utilizing balanced precision navigation in solving the item) across all 10 items, surpassing the proportions in the other two latent states for each group. For the easiest item (item 1), which requires a singular skill, all three latent profile groups exhibit the highest proportion of participants in State 3 (quick navigation).

Examining differences among the groups, we observe variations in the distribution of the three latent states discovered by eye-tracking across different item types. Specifically, for items 3 to 7, the low and high latent attribute profile groups share a similar proportion of State 1, both higher than the medium latent attribute profile group. For the most complex item (item 9), where objects are challenging to visualize, about 75% of participants in the low latent attribute profile group belong to State 1 (Precision Navigation), with 20% in State 3 (Quick Navigation). In contrast, only 20% of the other two groups are in State 1, with the majority in State 2. We conducted a chi-square test to examine the association between the distribution of the latent states and the latent profile groups for each item. The results revealed a statistically significant difference in the distribution of the three latent states across the three latent profile groups for item 9, $\chi^2(4) = 19$, $p < .001$. These findings underscore both between-group and within-group individual differences.

Figure 11. The first 10 state trajectories of the high latent profile group.

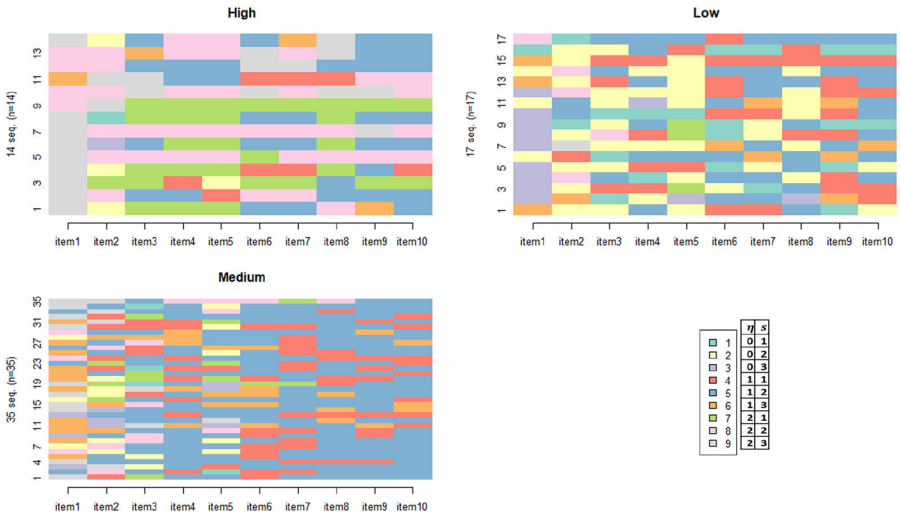

Figure 12. Latent trajectories with combined latent levels.

Figure 11 depicts the eye-tracking hidden state trajectories for the first 10 participants in the high latent attribute profile group. Despite all being classified as having high skill mastery by fluency CDM, they employ distinct eye-movement patterns, as evidenced by their membership in different latent states for different questions. Figure 12 combines item latent scores using levels from eye-tracking hidden states and ideal response scores for each participant. This reveals additional individual differences among participants.

Table 6
The Item Characteristics Based on Fluency CDM and Eye-Tracking Hiden States

Item	Measured Attributes	g	s_1	s_2	a	γ	State 1	State 2	State 3
1	x90	.884	.028	.012	1.09	2.515	.015	.273	.712
2	y90	.575	.164	.064	2.092	3.171	.197	.576	.227
3	x180	.518	.120	.049	2.172	3.512	.348	.561	.091
4	x90y180	.490	.189	.089	1.955	3.693	.303	.606	.091
5	y180	.326	.312	.152	2.246	3.652	.273	.606	.121
6	x90y90	.422	.202	.104	2.124	3.551	.379	.485	.136
7	x90y90	.321	.325	.164	1.950	3.635	.364	.576	.061
8	x90y180	.373	.115	.050	1.942	3.722	.227	.667	.106
9	x180y90	.295	.406	.230	2.013	3.757	.318	.561	.121
10	x180y90	.501	.133	.060	1.802	3.688	.258	.682	.061

Item Characteristics

The fluency CDM's item parameters and the hidden state distribution for each item are detailed in Table 6. These fluency CDM item parameters align with findings from a prior study that analyzed binary responses and response times using the same spatial rotation learning platform instrument (Zhang & Wang, 2018). In essence, items designed to measure a single skill exhibit relatively large guessing parameters (g), small time intensity parameters (γ), and minimal time discrimination parameters (a). Conversely, items assessing two or more complex skills display relatively large sleeping parameters (s_1, s_2), substantial time intensity parameters, and elevated time discrimination parameters. Beyond these established characteristics, the distribution of participants across three hidden states also provides additional insights into item characteristics. Notably, Item 1, serving as the initial question in the assessment and measuring a single, straightforward skill (rotation along the x-axis by 90), shows the highest proportion of participants in hidden state 3. As the complexity of skill or the number of skills for an item increases, there is a general decrease in the proportion of participants in state 3 and an increase in the proportion in State 1. This trend is evident in Figure 10.

Additionally, we computed the correlation between hidden state proportions and fluency CDM item parameters (g, s_1, s_2, a, γ). Employing a significance level of.05, we observed that item-state 3 proportion has a significantly positive association with the guessing parameter g ($r = .86$, $p = .035$) and a significantly negative association with the time intensity parameter ($r = -.96$, $p < .001$). The proportion of item-state 2 is significantly positively related to time intensity parameters ($r = .86$, $p = .035$). These findings suggest that items categorized as relatively easy are more likely to be addressed through eye-tracking movement reflected by state 3 (quick Navigation). Items that are more difficult require are likely to require balanced Navigation (state 2).

Figure 13 visually represents each item through Chernoff Faces, utilizing the 8 variables outlined in Table 6. It is easy to observe that Item 1 stands out distinctly from the other items. Items 5, 7, and 9 exhibit closer proximity to each other

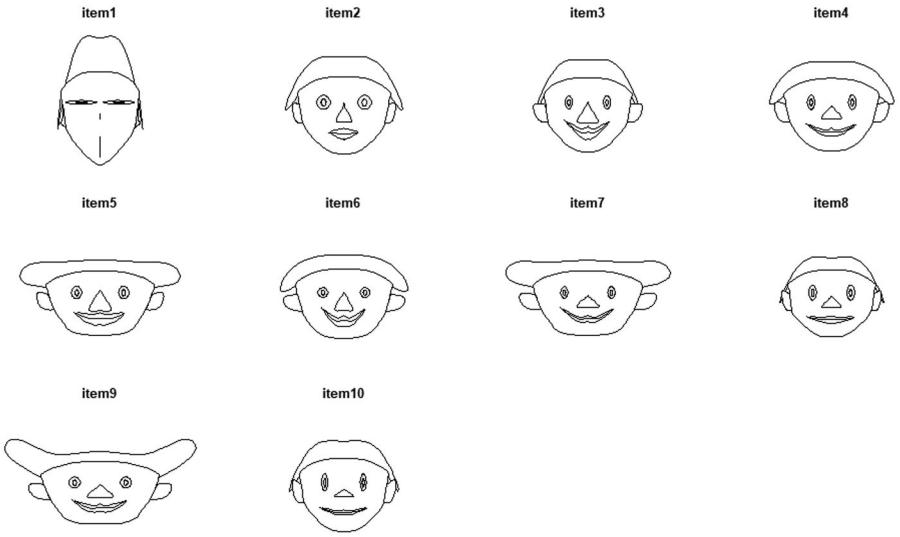

Figure 13. Items charted by both fluency CDM item parameters and eye tracking state distribution.

compared to the remaining items, while Items 8 and 10 display similarities. The remaining items—2, 3, 4, 6, and 8—show some distinct features from one another.

Discussion

This exploratory study presents a comprehensive multimodal data analysis framework that employs several statistical and machine learning methods. These methods aim to unveil cognitive outcomes and the problem-solving process by examining response accuracy, response times, and eye-tracking variables. Our findings related to the two primary research questions posed in "Introduction" section are summarized below.

The first research question explores the connection between hidden states identified through eye-tracking features and the latent attribute profiles estimated from the fluency CDM model based on response accuracy and response time. Our observations reveal that the empirical hidden state transition probability appears to be influenced by changes in latent attribute mastery levels between two consecutive items (refer to "Hidden States Distribution and Skill Mastery" section). This suggests the potential development of a new hidden Markov model that incorporates discrete latent attribute profiles as covariates. Such a model could better capture varying transition probabilities based on different levels of attribute skill mastery. This is similar to the idea from a recent study by Tang (2023) which proposed a latent hidden Markov model that integrates continuous latent traits. This model was utilized to describe the response process and its variations among respondents. Additionally, our investigation indicates that participants with the same or similar latent attribute profiles, as determined by the fluency CDM, exhibit different hidden state trajectories based on their eye-tracking features. This suggests that eye-tracking features can offer addi-

tional diagnostic insights into the problem-solving processes related to spatial rotation questions. Figure 12 illustrates how a combined representation of latent levels can describe more individual differences.

The second research question aims to identify additional insights that can be gleaned from the eye-tracking features related to the spatial rotation items. The findings, as detailed in "Item Characteristics" section, demonstrate that the distribution of participants across various hidden states reflects the complexity of visualizing objects in each item. This introduces an additional dimension to the characterization of item features, complementing the information obtained from the item parameters in the fluency CDM model, which is based on response accuracy and response time.

The limitations of the current study can be succinctly outlined as follows. First, the generalizability of the findings may be constrained to context-specific scenarios or relatively small sample sizes. Studies utilizing eye-tracking variables typically feature small sample sizes, ranging from 26 to 90 participants (Hu et al., 2017; Kaczorowska et al., 2021; Yaneva et al., 2021; Zhu & Feng, 2015; Zhan et al., 2022). Our study's valid sample size of 70 falls within this established range. The smaller sample size is primarily attributed to the limited availability of eye-trackers during the experiment, which significantly hindered the efficient collection of eye-tracking data. In the current study, the stability of the feature selection procedure is constrained by the small sample size. Consequently, further validation of the selected features is warranted through future studies with larger sample sizes. Another potential factor that impacts the findings of the current study is the method of defining an Area of Interest (AOI). Since we rely on numerous eye-tracking variables aggregated over a specified AOI, the definition of this area significantly influences the information summarized from it. In our current study, we established three reasonable AOIs based on the item's design to assess spatial rotation abilities. However, these AOIs could be further segmented into smaller, more specific regions to capture more detailed information about eye movement. For instance, defining an AOI based on a specific side of an object could offer a more granular perspective. The final note is that our current study still utilizes summary variables from the eye-tracker. In future research, it may be worthwhile to directly analyze eye-movement action sequences, describing how a participant moves within several defined AOIs across the entire screen. This approach goes beyond utilizing measures within a singular AOI and provides a more comprehensive understanding of eye-movement patterns.

Appendix

The appendix shows the details of data processing procedures, model parameters, and more experimental results.

Appendix 1: Procedure of Data Preprocessing

Step	Description	Variables Left
1	Remove redundant and meaningless information.	82
2	Remove variables with too many missing values.	65

Variables Removed due to Missing Values

Table A1
Variable Names and Their Units

Variable Name	Unit
Standard deviation of peak velocity of saccades	degree/ms
Time to exit saccade	ms
Time to entry saccade	ms
Time to exit saccade (look)	ms
Time to entry saccade (look)	ms
Time to exit saccade (option)	ms
Time to entry saccade (option)	ms
Peak velocity of exit saccade	degree/ms
Time to exit saccade (question)	ms
Peak velocity of entry saccade	degree/ms
Time to entry saccade (question)	ms
Peak velocity of exit saccade (look)	degree/ms
Peak velocity of entry saccade (look)	degree/ms
Peak velocity of exit saccade (option)	degree/ms
Peak velocity of entry saccade (option)	degree/ms
Peak velocity of exit saccade (question)	degree/ms
Peak velocity of entry saccade (question)	degree/ms

Variables Removed related to "Time to First"

Table A2
Variable Names and Their Units

Variable Name	Unit
Time to first fixation	ms
Time to first fixation (look)	ms
Time to first fixation (option)	ms
Time to first fixation (question)	ms
Time to first Glance	ms
Time to first Glance (look)	ms
Time to first Glance (option)	ms
Time to first Glance (question)	ms
Time to first Visit	ms
Time to first Visit (look)	ms
Time to first Visit (option)	ms
Time to first Visit (question)	ms
Time to first saccade	ms

Variables Removed related to "Visits"

Table A3
Variables and Their Units

Variable Name	Unit
Number of Visits (look)	count
Duration of first Visit	ms
Number of Visits (option)	count
Total duration of Visit	ms
Time to first Visit (look)	ms
Average duration of Visit	ms
Maximum duration of Visit	ms
Minimum duration of Visit	ms
Time to first Visit (option)	ms
Duration of first Visit (look)	ms
Time to first Visit (question)	ms
Total duration of Visit (look)	ms
Average duration of Visit (look)	ms
Duration of first Visit (option)	ms
Maximum duration of Visit (look)	ms
Minimum duration of Visit (look)	ms
Total duration of Visit (option)	ms
Average duration of Visit (option)	ms
Duration of first Visit (question)	ms
Maximum duration of Visit (option)	ms
Minimum duration of Visit (option)	ms
Total duration of Visit (question)	ms
Average duration of Visit (question)	ms
Maximum duration of Visit (question)	ms
Minimum duration of Visit (question)	ms

Full List of Variables after Data Preprocessing

After removing the variables with large proportion of missing values, and some variable contains meaningless information, there are 65 variables left. Here we give the full list of the variables after data preprocessing, including their units.

Table A4
Variables and Their Units

Variable Name	Unit
Pupil	mm
Number of fixations	count
Number of fixations (look)	count
Duration of first fixation	ms
Number of fixations (option)	count

(*Continued*)

Table A4
(Continued)

Variable Name	Unit
Total duration of fixations	ms
Number of fixations (question)	count
Average duration of fixations	ms
Maximum duration of fixations	ms
Minimum duration of fixations	ms
Duration of first fixation (look)	ms
Total duration of fixations (look)	ms
Duration of first fixation (option)	ms
Average duration of fixations (look)	ms
Maximum duration of fixations (look)	ms
Minimum duration of fixations (look)	ms
Total duration of fixations (option)	ms
Duration of first fixation (question)	ms
Average duration of fixations (option)	ms
Maximum duration of fixations (option)	ms
Minimum duration of fixations (option)	ms
Total duration of fixations (question)	ms
Average duration of fixations (question)	ms
Maximum duration of fixations (question)	ms
Minimum duration of fixations (question)	ms
Number of Glances	count
Number of Glances (look)	count
Duration of first Glance	ms
Number of Glances (option)	count
Total duration of Glances	ms
Number of Glances (question)	count
Average duration of Glances	ms
Maximum duration of Glances	ms
Minimum duration of Glances	ms
Duration of first Glance (look)	ms
Total duration of Glances (look)	ms
Duration of first Glance (option)	ms
Average duration of Glances (look)	ms
Maximum duration of Glances (look)	ms
Minimum duration of Glances (look)	ms
Total duration of Glances (option)	ms
Duration of first Glance (question)	ms
Average duration of Glances (option)	ms
Maximum duration of Glances (option)	ms
Minimum duration of Glances (option)	ms
Total duration of Glances (question)	ms
Average duration of Glances (question)	ms
Maximum duration of Glances (question)	ms
Minimum duration of Glances (question)	ms
Number of saccades	count

(*Continued*)

Table A4
(Continued)

Variable Name	Unit
Average peak velocity of saccades	degree/ms
Minimum peak velocity of saccades	degree/ms
Maximum peak velocity of saccades	degree/ms
Average amplitude of saccades	degree
Minimum amplitude of saccades	degree
Maximum amplitude of saccades	degree
Total amplitude of saccades	degree
Direction of first saccade	degree
Peak velocity of first saccade	degree/ms
Average velocity of first saccade	degree/ms
Amplitude of first saccade	degree
Number of saccades in AOI	count
Number of saccades in AOI (look)	count
Number of saccades in AOI (option)	count
Number of saccades in AOI (question)	count

Appendix 2: The Three-Step Feature Selection Results

Variables Removed in the First Step: Accuracy as Predictions

Table A1
Variables and Their Units

Variable Name	Unit
Number of fixations	count
Number of fixations (look)	count
Number of fixations (option)	count
Number of fixations (question)	count
Total duration of fixations (question)	ms
Number of saccades in AOI	count

Variables Removed in the First Step: Response Time as Predictions

Table A2
Variables and Their Units

Variable Name	Unit
Number of fixations	count
Number of fixations (look)	count
Number of fixations (option)	count
Number of fixations (question)	count
Total duration of fixations (question)	ms
Number of saccades in AOI	count

Variables Selected in the Second Step: Accuracy as Predictions

Table A3
Variables and Their Units

Variable Name	Unit
Total duration of fixations	ms
Average duration of fixations (option)	ms
Average duration of Glances (option)	ms
Minimum duration of Glances (option)	ms
Total duration of Glances (question)	ms
Maximum duration of Glances (question)	ms
Maximum peak velocity of saccades	degree/ms

Variables Selected in the Second Step: Response Time as Predictions

Table A4
Variables and Their Units

Variable Name	Unit
Number of fixations	count
Number of fixations (look)	count
Number of fixations (question)	count
Maximum duration of fixations (look)	ms
Maximum duration of fixations (question)	ms
Minimum duration of fixations (question)	ms
Number of Glances	count
Total duration of Glances	ms
Average duration of Glances	ms
Maximum duration of Glances	ms
Minimum duration of Glances	ms
Average duration of Glances (option)	ms
Minimum duration of Glances (option)	ms
Total duration of Glances (question)	ms

Note

[1]The total sample size is 66 due to 4 participants have some missing values on some of the selected eye-tracking features, thus do not have hidden states estimated from the HMM

References

Baker, F. B. (2001). *The basics of item response theory*. ERIC.

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4: Linear mixed-effects models using eigen and s4. r package version 1.1-7.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472. doi: 10.1214/ss/1177011136.

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology*, *61*(5), 683–689.

Lab, T. (2023). Understanding Tobii Pro Lab's eye tracking metrics. Retrieved from https://connect.tobii.com/s/article/understanding-tobii-pro-lab-eye-tracking-metrics? Accessed: 2023-06-19.

Hu, Y., Wu, B., & Gu, X. (2017). An eye tracking study of high-and low-performing students in solving interactive and analytical problems. *Journal of Educational Technology & Society*, *20*(4), 300–311.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480.

Kaczorowska, M., Plechawska-Wójcik, M., & Tokovarov, M. (2021). Interpretable machine learning models for three-way classification of cognitive workload levels for eye-tracking features. *Brain Sciences*, *11*(2), 210.

Khooshabeh, P., & Hegarty, M. (2010). Representations of shape during mental rotation. In *2010 AAAI spring symposium series*. Association for the Advancement of Artificial Intelligence.

Man, K., & Harring, J. R. (2023). Detecting preknowledge cheating via innovative measures: A mixture hierarchical model for jointly modeling item responses, response times, and visual fixation counts. *Educational and Psychological Measurement*, *83*(5), 1059–1080.

Nazareth, A., Killick, R., Dick, A. S., & Pruden, S. M. (2019). Strategy selection versus flexibility: Using eye-trackers to investigate strategy use during mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 232.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Tang, X. (2023). A latent hidden Markov model for process data. *Psychometrika*, *89*(1), 205–240.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181–204. doi: 10.3102/10769986031002181.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*(3), 287–308. doi: 10.1007/s11336-006-1478-z.

Visser, I., & Speekenbrink, M. (2010). depmixs4: An r package for hidden Markov models. *Journal of Statistical Software*, *36*, 1–21.

Wang, S., & Chen, Y. (2020). Using response times and response accuracy to measure fluency within cognitive diagnosis models. *Psychometrika*, *85*(3), 600–629.

Wang, S., Hu, Y., Wang, Q., Wu, B., Shen, Y., & Carr, M. (2020). The development of a multidimensional diagnostic assessment with learning tools to improve 3-d mental rotation skills. *Frontiers in Psychology*, *11*, 305.

Wang, S., Zhang, S., & Shen, Y. (2020). A joint modeling framework of responses and response times to assess learning outcomes. *Multivariate Behavioral Research*, *55*(1), 49–68.

Wang, Z., Tang, X., Liu, J., & Ying, Z. (2023). Subtask analysis of process data through a predictive model. *British Journal of Mathematical and Statistical Psychology*, *76*(1), 211–235.

Weisberg, S. (2005). *Applied linear regression* (Vol. *528*). John Wiley & Sons.

Xiao, Y., He, Q., Veldkamp, B., & Liu, H. (2021). Exploring latent states of problem-solving competence using hidden Markov model on process data. *Journal of Computer Assisted Learning*, *37*(5), 1232–1247.

Xue, J., Li, C., Quan, C., Lu, Y., Yue, J., & Zhang, C. (2017). Uncovering the cognitive processes underlying mental rotation: An eye-movement study. *Scientific Reports*, *7*(1), 10076.

Yaneva, V., Clauser, B. E., Morales, A., & Paniagua, M. (2021). Using eye-tracking data as part of the validity argument for multiple-choice questions: A demonstration. *Journal of Educational Measurement*, *58*(4), 515–537.

Yaneva, V., Clauser, B. E., Morales, A., & Paniagua, M. (2022). Assessing the validity of test scores using response process data from an eye-tracking study: A new approach. *Advances in Health Sciences Education*, *27*(5), 1401–1422.

Zhan, P., Man, K., Wind, S. A., & Malone, J. (2022). Cognitive diagnosis modeling incorporating response times and fixation counts: Providing comprehensive feedback and accurate diagnosis. *Journal of Educational and Behavioral Statistics*, *47*(6), 736–776.

Zhang, S., & Wang, S. (2018). Modelling learner heterogeneity: A mixture learning model with responses and response times. *Frontiers in Psychology*, *9*, 2339.

Zhu, M., & Feng, G. (2015). An exploratory study using social network analysis to model eye movements in mathematics problem solving. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 383–387). Association for Computing Machinery.

Authors

SHIYU WANG is associate professor, Quantitative Methodology Program, Department of Educational Psychology, University of Georgia, 325N Aderhold Hall, Athens, GA, 30602; swang44@uga.edu. Her primary research interests encompass computerized adaptive testing and innovations in latent variable modeling.

SHUSHAN WU is a PhD candidate, Department of Statistics, University of Georgia, 310 Herty Drive, Athens, Georgia, GA, 30605; Shushan.Wu@uga.edu. Her primary research interests focus on geometric machine learning, network data analysis, big data analysis, and applications in educational measurement.

YINGHAN CHEN is associate professor, Department of Mathematics and Statistics, University of Neveda, Reno, 1664N. Virgina Street, Reno, NV 89557; yinghanc@unr.edu; Her primary research interests include Bayesian analysis, computational statistics and latent variable models.

LUYANG FANG is a PhD candidate, Department of Statistics, University of Georgia, 310 Herty Drive, Athens, Georgia, GA 30605; Luyang.Fang@uga.edu. Her primary research interests focus on big data analysis, nonparametric modeling, deep learning and LLM.

LIANG XIAO is a graduate student, the College of Education, Zhejiang Normal University, Jinhua 321001, China; sherlockvcm@163.com. His primary research interests lie in the areas of educational technology, learning analytics.

FEIMING LI is professor, the College of Education, Zhejiang Normal University, Jinhua 321001, China; feimingli@zjnu.edu.cn. Her primary research interests lie in the areas of psychometrics, technology-enhanced assessment, and learning analytics.