Analysis of Problem-Solving Process Strategies in Computerized Dynamic Testing: Extension and Application of Multilevel Mixture IRT Models

Authors: Li Meijuan, Liu Yue, Liu Hongyun, Liu Hongyun

Date: 2019-11-08T00:00:00+00:00

Abstract

Students generate extensive time-stamped process data during computer-based dynamic assessments. Based on 139,990 data points from 3,196 students across five countries/regions on a traffic problem-solving task from PISA 2012, this study extends the multilevel mixture IRT (MMixIRT) model to explore the categorical characteristics of problem-solving process strategies. Results demonstrate that the model can not only analyze typical features of strategy usage among students from different countries/regions during problem-solving based on behavioral sequences, but also provide individual-level ability estimates. The extended MMixIRT model can be utilized to analyze characteristics of process data.

Full Text

Analysis of Problem-Solving Strategies in Computer-Based Dynamic Assessment: Extension and Application of the Multilevel Mixture IRT Model

Meijuan Li[1], , Yue Liu[2], Hongyun Liu[2,3]

[1]Educational Supervision and Quality Assessment Research Center, Beijing Academy of Educational Sciences, Beijing 100036, China
[2]Faculty of Psychology, Beijing Normal University, Beijing 100875, China
[3]Beijing Key Laboratory of Applied Experimental Psychology, Faculty of Psychology, Beijing Normal University, Beijing 100875, China
Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing 100875, China

Abstract

[**Objective**] This research proposes an extended Multilevel Mixture IRT (MMixIRT) model to detect and interpret problem-solving strategies based on process data. [**Methods**] The extended model incorporates two key modifications: first, it uses accumulated information from steps as process data for specific steps; second, it defines a more flexible design matrix to determine the information used for ability estimation at the individual level. [**Results**] The extended model can analyze typical characteristics of strategy usage among students from different countries/regions at the process level while simultaneously providing ability estimates at the individual level. [**Limitations**] The model' s application requires recoding process data into a structure similar to that used in this study, which may not be feasible for all tasks in practice. [**Conclusions**] The extended model is suitable for analyzing features of process data.

Keywords: Computer-based dynamic assessment; Problem-solving strategy; Process data; Extended MMixIRT model
Classification Code: B841

Problem-solving ability refers to the capacity to understand and resolve problems in situations where solutions are not immediately apparent through a series of cognitive processes (Mayer, 1982) [?]. During this process, problem solvers must fully comprehend the core of the problem, design feasible solutions, implement them, and control progress to achieve goals (Garofalo & Lester, 1985) [?]. Problem-solving ability is crucial for learning and success, and many large-scale international educational assessment programs have prioritized its evaluation, such as the Programme for International Student Assessment (PISA) (OECD, 2003, 2013) [?][?]. In recent years, advances in information technology and research in computer-based testing have provided new approaches for assessing problem-solving ability. For instance, PISA 2012 employed computer-based dynamic assessment to evaluate students' problem-solving abilities by simulating real-life scenarios, focusing on how students apply general cognitive processes in the absence of explicit solutions (OECD, 2013) [?] and emphasizing the dynamic and interactive nature of the problem-solving process (Funke, 2001) [?].

Computer-based testing can transform not only test design and administration but also data analysis methods (DiCerbo & Behrens, 2012) [?]. Such systems can examine whether students answer correctly and automatically record time-stamped behavioral sequences (Kerr, Chung, & Iseli, 2011) [?], capturing the time students spend and their series of behaviors during problem solving—data known as process data (Zoanetti, 2010) [?]. Based on process data, researchers can analyze and mine students' problem-solving strategies while also using this information as evidence for problem-solving ability evaluation (DiCerbo & Behrens, 2012) [?]. For example, Greiff, Wüstenberg, and Avvisati (2015) [?] found, based on PISA 2012 process data from the "Room Temperature Control" task, that the strategy of changing only one variable at a time could predict both performance on that specific item and overall problem-solving scores. In recent years, with developments in measurement theory and statistical techniques, increasing attention has been paid to problem-solving processes, skills, and strategies. One approach involves identifying the skills (or attributes) required for items and analyzing problem-solving strategy characteristics based on measurement models, with cognitive diagnosis models being particularly representative. For instance, de la Torre and Douglas (2004) [?] used higher-order latent structure models to estimate student abilities and classify students' cognitive characteristics based on their cognitive attribute mastery patterns. Another approach employs statistical models and data mining techniques to analyze the rich information embedded in process data, including visualization analysis (DiCerbo, Liu, Rutstein, Choi, & Behrens, 2011) [?], cluster analysis (Bergner, Shu, & von Davier, 2014) [?], and classification analysis (Desmarais & Baker, 2012) [?]. Recently, scholars (Shu, Bergner, Zhu, Hao, & von Davier, 2017) [?] have combined Hidden Markov Models with item response models to analyze sequential response information in process data for ability estimation. The method explored in this study belongs to the second category: analyzing different strategies students employ during problem solving based on process data while simultaneously estimating ability based on task submission status information.

Process data has a nested structure, with behavioral sequences generated during task completion (i.e., process-level data) nested within individual students. Therefore, multilevel framework models can be adapted for process data analysis (Goldstein, 1987) [?]. The Multilevel Mixture Item Response Theory (MMixIRT) model combines multilevel modeling with mixture IRT, improving parameter estimation precision while obtaining measurement characteristics of different latent class groups (Cho & Cohen, 2010) [?]. For two-level data, MMixIRT can analyze discontinuous latent variables (latent classes) and continuous latent variables (ability) at both Level 1 and Level 2. Level 1 latent class analysis is primarily based on relationships among respondents' item responses, while Level 2 latent class analysis focuses on relationships among group members' response patterns (Vermunt, 2003) [?]. Although MMixIRT provides a framework for analyzing nested data and categorical features, direct application to process data presents two problems: (1) A single step in the process reflects only one operation or behavioral performance at a specific time point, violating the model' s assumption that measurements at different time points represent manifestations of a trait at each moment. (2) Using all process data to estimate individual ability introduces bias and interpretation difficulties due to inconsistent traits measured across different problem-solving stages or steps. Consequently, traditional MMixIRT models are not suitable for process data in terms of model assumptions and interpretation of latent variable meaning. The challenge lies in adapting this modeling approach for process data.

Internationally, increasing research has focused on mining process data to analyze typical characteristics of problem solving across different student groups (Qiao & Jiao, 2018; Liao, He, & Jiao, 2019) [?][?]. However, most studies have used only partial response information or focused solely on categories while ignoring ability estimation. Few studies have simultaneously examined problem-solving strategy categories and individual-level information reflecting problem-solving ability based on the nested nature of process data. Using a PISA 2012 problem-solving item as an example, this study extends the MMixIRT model based on process data from five countries/regions (hereinafter referred to as regions) to analyze different strategies students employ during problem solving, estimate individual-level abilities, and compare strategy usage characteristics across regions.

2. The Extended MMixIRT Model

The definition and detailed description of the traditional MMixIRT model can be found in Cho and Cohen (2010) [?]. This study modifies and extends the traditional MMixIRT model in two aspects.

First, to capture the continuous nature of behavioral sequences in problem-solving tasks, accumulated step information is used as process data for specific steps, expressed as:

$$Y_{jki} = \sum w_t y_{tki} \quad (1)$$

where y_{tki} represents the operational behavior of student k at time point t on scoring point i (similar to paths in the subsequent traffic problem). While the traditional MMixIRT models y_{tki} directly, the extended MMixIRT models the accumulated response Y_{jki}. If time $t = j$, $w_t = 1$; otherwise $w_t = 0$, which reduces to the traditional MMixIRT model. Based on test item and process data characteristics, this study uses cumulative responses as the process j response, i.e., $w_t = 1$ if $t \leq j$.

Second, to enable more flexible decomposition of variance at process and individual levels, design matrix $\mathbf{A}$ is defined to decompose variation across process and individual levels, where row j, $\mathbf{A}_j$, defines the decomposition weights for latent variables at different levels of process data. The extended model can be expressed as:

$$Y_{jki} = \mathbf{A}_j(\cdot) \quad (2)$$

In traditional MMixIRT models, $Y_{jki} = Y_{jki}^{(w)} + Y_{ki}^{(B)}$, decomposing Y_{jki} variation into Level 1 (within-group $Y_{jki}^{(w)}$) and Level 2 (between-group $Y_{ki}^{(B)}$) components. If for any j, design matrix $\mathbf{A}_j = (1, 1)$, this becomes the traditional MMixIRT model. Thus, the traditional model is a special case of the extended model. The differences between extended and traditional models manifest in two aspects: (1) The latent class at each process-level step represents the cumulative state of previous steps rather than performance at that single step, which better explains problem-solving strategy usage and provides a basis for exploring strategy continuity and transitions. (2) The measurement indicators used to define individual-level latent variables differ from traditional MMixIRT models. Traditional models estimate individual-level latent variables from Level 1 observed variables $[y_{jk1}, \dots, y_{jki}, \dots, y_{jkI}]$ (Lee, Cho, & Sterba, 2017) [?], whereas the extended model can define a more flexible design matrix $\mathbf{A}$ to determine information used for individual-level ability estimation.

3. The Extended MMixIRT Model Used in This Study

The extended MMixIRT model is flexible, allowing different models to be defined at Level 1 and Level 2 based on research foci. Considering process data characteristics, this study focuses on differences in problem-solving strategies and individual ability reflected in final status. Therefore, the model used is also a special case of the aforementioned extended model.

3.1 Model Definition

The extended MMixIRT model used in this study comprises two levels: process level and individual level. At the process level, latent classes are defined to describe heterogeneity across steps for strategy classification. At the individual level, continuous latent variables estimate individual ability.

Process-level model:

$$P(Y_{jk1} = S_1, \dots, Y_{jkI} = S_I) = \sum P(C_{jk} = g) P(Y_{jk1} = S_1, \dots, Y_{jkI} = S_I | C_{jk} = g)$$

$P(Y_{jk1} = S_1, \dots, Y_{jkI} = S_I)$ represents the probability that after step j ($j = 1, \dots, J_k$, where J_k denotes the total number of steps for student k), the response status on scoring points is $(S_1, \dots, S_I)$ (note that the number of steps J_k varies across students). $P(C_{jk} = g)$ represents the probability that step j for student k belongs to latent class g ($g = 1, 2, \dots, G$), where G is the number of latent classes. $P(Y_{jk1} = S_1, \dots, Y_{jkI} = S_I | C_{jk} = g)$ represents the conditional probability that the accumulated response status for the first j steps is $(S_1, \dots, S_I)$ given that step j for student k belongs to latent class g.

Individual-level model:

$$P(y_{ki} = 1 | \theta_k) = \frac{\exp(\alpha_i \theta_k - \beta_i)}{1 + \exp(\alpha_i \theta_k - \beta_i)}$$

The individual-level model estimates individual ability based on students' final response status. The corresponding design matrix $\mathbf{A}$ is defined as: $\mathbf{A}_j = (1, 1)$ if j is the student' s final submission step, otherwise $\mathbf{A}_j = (1, 0)$. In the individual-level model, y_{ki} represents student k's response on scoring point i. α_i denotes the discrimination parameter for scoring point i, β_i denotes the difficulty parameter for scoring point i ($i = 1, 2, \ldots, I$), and θ_k represents the ability estimate for student k based on the final step in the process. θ_k is assumed to follow a standard normal distribution ($\theta_k \sim N(0, 1)$).

[Figure 1: see original paper] illustrates the basic structure of the extended MMixIRT model used in this study. Squares represent students' response reactions during the process, circles represent latent variables, and the triangle containing 1 represents a constant vector of ones (whose coefficients correspond to intercept parameters β_i, i.e., difficulty parameters in traditional IRT models). At the process level, C_{jk} is a categorical latent variable, while at the individual level, θ_k is a continuous latent variable. At the process level, student k' s responses to all paths at step j, $[y_{jk1}, \ldots, y_{jki}, \ldots, y_{jkI}]$, can be explained by categorical latent variable C_{jk}. At the individual level, student k' s final responses to all paths, $[y_{k1}, \ldots, y_{ki}, \ldots, y_{kI}]$, can be explained by continuous latent variable θ_k. According to Equation (4), arrows from continuous latent variable θ_k to each path' s response state describe how changes in ability θ_k affect the probability of selecting that path, corresponding to discrimination parameters (α_i). Arrows from the triangle to each path represent the probability of selecting that path when $\theta_k = 0$, corresponding to difficulty parameters (β_i) in traditional IRT models.

3.2 Parameter Recovery and Classification Accuracy

A Monte Carlo simulation study was conducted to examine parameter recovery and classification accuracy of the proposed model. The design considered two factors: (1) number of latent classes at the process level (3 and 5), and (2) number of process steps completed by individuals (30 and 50 steps), resulting in $2 \times 2 = 4$ experimental conditions. Using self-written R programs, response data were generated for each condition based on the extended MMixIRT model, with $\alpha_i \sim U(1, 2.5)$, $\beta_i \sim N(0, 1)$, and $\theta_k \sim N(0, 1)$ (Wang, Xu, Shang, & Kuncel, 2018) [**?**]. Response probabilities for different classes followed Nylund, Asparouhov, and Muthén (2007) [**?**]. True values for class proportions and item (path) correct response probabilities under different conditions are provided in Appendix 1. Each condition assumed equal process steps across all individuals, with the final step representing each individual' s final response status for ability estimation. Sample size was fixed at 600 respondents per condition, with 100 replications. Model parameters were estimated using Mplus 7.11 (Muthén & Muthén, 2005) [**?**].

Results indicated good parameter recovery, evidenced by small biases across all parameters. Root mean square error (RMSE) was approximately 0.2 for discrimination parameters, below 0.1 for difficulty parameters, and around 0.3 for ability parameters. Classification accuracy was high across all conditions, exceeding 96%.

4.1 Task

This study used a traffic problem item from the PISA 2012 problem-solving assessment (Traffic CP007Q02). The map indicated the time required for each path, requiring students to find the fastest route from Diamond to Einstein. The correct shortest path required 31 minutes. Task description and path identification are shown in [Figure 2: see original paper].

Note: Routes between two nodes on the map represent paths. The blue-highlighted path indicates the correct route. [Figure 2: see original paper] PISA 2012 Traffic Problem Task and Correct Path

4.2 Process Data Coding

The process data were obtained from http://www.oecd.org/pisa/data/. First, information related to valid path clicks was filtered. Then, "path selection situations" were split by different paths to obtain click results for 23 paths (P1, P2, P3, ···, P23). presents the formatted data structure, where each row represents a step in a student' s response process and each column represents a path. Here, 0 indicates not selected and 1 indicates selected. For example, the first row shows student ID 00017 selecting P2 at Step 1, the second row shows selecting P1 at Step 2, the third row shows selecting P13 at Step 3, and the eighth row shows canceling P1 at Step 8.

Subsequently, responses were rescored based on correctness. Similar to traditional test analysis, variables P1, P2, P3, ···, P23 in represent 23 paths. The correct path is: Diamond-Nowhere-Sakharov-Market-Lee-Mandela-Einstein, corresponding to P1, P5, P7, P8, P13, and P17. For each step in the process, if a student selected a correct path, that path was scored 1; otherwise, 0. Similarly, if an incorrect path was selected, that path was scored 0; otherwise, 1. The recoded 23 variables were named "CP1, CP2, CP3, ···, CP23." presents examples of the recoded data format. For instance, the first row shows student ID 00017 selecting P2 (an incorrect path) at Step 1; the second row shows selecting P1 (a correct path) at Step 2, etc.

4.3 Sample

The study sample comprised 3,196 15-year-old students from five regions in the PISA 2012 problem-solving assessment. Sample sizes were 1,449 for Canada, 433 for Hong Kong-China, 411 for Shanghai-China, 456 for Singapore, and 406 for the United States. The five regions generated 139,990 process steps, with an average of 43.80 steps per student (SD = 38.06), ranging from 1 to 335 steps. Mean response time was 669.22 seconds (SD = 543.12 seconds), ranging from 10.7 to 2,384.7 seconds.

4.4 Data Analysis Methods

The extended MMixIRT model was employed to estimate strategy categories and individual abilities using Mplus 7.11. Association rule mining was used to explore relationships among different strategy categories.

Association rule mining aims to identify correlations among items appearing simultaneously in large datasets. If regular patterns exist between two or more variables, they are considered associated. The Apriori algorithm is a commonly used frequent itemset mining algorithm for association rules, which recursively generates frequent itemsets starting from single-item sets, then two-item sets, and continues recursively until all frequent itemsets are identified (Peter, 2013) [?]. This study used the Apriori algorithm on the SPMF platform to further analyze relationships among students' problem-solving strategies.

4.5 Variables

Three process-related variables (number of path clicks, number of resets, and response time) were used to validate model estimation results through correlation analysis. Number of path clicks indicates how many paths students clicked; number of resets indicates how many times students canceled all previous path selections and restarted; response time indicates the time taken to complete the task. The study also included the absolute difference between time spent and correct response time, representing the deviation between the final submission path' s time and the correct answer time (31 minutes).

5.1 Model Selection

For extended MMixIRT model analysis, the number of latent classes must first be determined based on model fit indices and interpretability of latent classes (Rosato & Baer, 2012) [?]. presents model fit indices for 1 to 7 classes estimated simultaneously across five regions. Fit indices include loglikelihood, AIC (Akaike, 1974) [?], BIC (Schwarz, 1978) [?], aBIC (Tofighi & Enders, 2008) [?], and entropy (Asparouhov & Muthén, 2014) [?]. The first four indices indicate better fit when smaller, while entropy measures the degree of separation among latent classes in mixture models, with values closer to 1 indicating better class separation. Results show that more latent classes yield better model fit. However, with 7 classes, two classes' paths could not form complete routes from start to end. With 6 classes, one class's paths could not form a complete route. Therefore, considering both fit indices and class interpretability, the 5-class solution was selected.

5.2 Strategy Category Characteristics and Relationship with Individual Ability Levels

The extended MMixIRT model classifies students' post-operation status at each step into five categories. The number of clicks on each path by latent class is provided in Appendix 2. Analyzing the most frequently selected paths and their associations within each class reveals typical paths for each category. [Figure 3: see original paper] shows the typical routes and sequences selected by each class, where circled numbers indicate path order and each class represents a problem-solving strategy.

Because students' operational behaviors are interrelated, the categories assigned to each behavior are also connected. If adjacent steps belong to different categories, students have changed strategies, indicating a strategy transition. Students' final strategy is highly correlated with ability values: using correct strategies leads to correct answers, while using incorrect strategies leads to incorrect answers. However, different types of incorrect strategies correspond to different ability levels.

Given the existence of strategy transitions, we used each student' s final step strategy as their ultimate strategy and analyzed mean ability estimates across strategies, yielding average ability values of -0.714, -1.281, -0.714, 0.399, and -0.714 for the five strategies. Combined with [Figure 3: see original paper], Strategy 4 matches the correct path with a time of 31 minutes and shows the highest individual ability values, indicating it is the correct strategy. Strategy 2 selects the longest route with no overlap with the correct path, requiring 35 or 36 minutes—the largest deviation from correct response time—and shows the lowest individual ability values, indicating it is the poorest strategy. Strategies 1, 3, and 5 partially overlap with the correct path. Although these strategies involve different paths and error types, their individual ability values are equal, suggesting minimal differences in quality among these strategies.

presents the distribution of strategies used by students from each region in their final step. The proportion of students using Strategy 4 (correct path) in their final step is highest, while the proportion using Strategy 2 (lowest ability) is lowest. Across regions, Singapore shows the highest proportion of students using Strategy 4 in their final step (81.6%), slightly higher than other regions, indicating Singaporean students performed best on this item. The United States shows the lowest proportion using Strategy 4 (75.6%) and the highest proportion using Strategy 2, indicating relatively poorer performance, consistent with individual ability level estimates. Additionally, error groups across regions show distinct characteristics in final-step strategy usage. For example, Canadian students more frequently used Strategy 3, Singaporean and Shanghai-Chinese students more frequently used Strategies 5 and 3, U.S. students more frequently used Strategy 1, and Hong Kong-Chinese students more frequently used Strategies 1 and 3.

5.3 Strategy Application

To explore strategy transitions in process data, this study defined obvious use of a typical strategy as continuous use of that strategy three or more times. In describing strategy transitions, only transitions between different strategies were recorded; if the same strategy appeared multiple times during a transition, only the final transition was recorded. presents the number of strategies applied by correct and error groups across regions. Overall, correct-group students most frequently applied 4 or 5 strategies. The proportion of correct-group students applying 5 strategies was significantly larger than the error group, indicating that nearly one-third of correct-group students found the correct route only after trying all 5 strategies. Over one-third of error-group students stopped after trying 4 strategies and submitted incorrect routes. Across regions, Singapore and the United States showed lower proportions of correct-group students applying 5 strategies, with the United States showing the lowest proportion.

and present association analysis results from the Apriori algorithm. Frequent itemsets refer to strategies that frequently co-occur, with frequency indicating how often two strategies appear together. Confidence refers to the proportion of transactions containing both antecedent and consequent among all transactions containing the antecedent. For example, 5==>1 indicates the proportion of students using both Strategy 1 and Strategy 5 among all students using Strategy 5. Results show that for correct-group students, Strategies 3 and 5, Strategies 3 and 4, Strategies 2 and 5, and Strategies 1 and 5 have strong associations. For instance, among students using Strategy 5, the probability of simultaneously using Strategy 3 is 0.51. In contrast, error-group students show weaker association between Strategies 3 and 4, indicating low probability of transitioning from Strategy 3 to the correct Strategy 4. Additionally, among students using Strategy 5, the probability of using Strategy 3 is significantly lower than in the correct group. shows that strategy usage rules are generally consistent across regions for both correct and error groups, except for Shanghai-China’ s correct group, which shows distinct patterns: probabilities of using Strategy 5 after Strategy 2 and using Strategy 5 after Strategy 3 are significantly higher than other regions, while probabilities of using Strategy 3 after Strategy 4 and using Strategy 4 after Strategy 3 are significantly lower.

5.4 Relationship Between Process Variables and Ability

presents descriptive statistics for process variables (number of path clicks, number of resets, difference between time spent and correct response time, response time) and their correlations with individual ability estimates. For all regions, correct groups show fewer path clicks and fewer resets than error groups, with no significant differences in response time between groups. Larger differences between time spent and correct response time correspond to lower average individual ability estimates. Results also reflect regional characteristics during task completion. For example, U.S. students show the lowest average individual ability estimates and fewest path clicks, with error groups showing significantly larger differences between time spent and correct response time than other regions. Singaporean students show the highest average individual ability estimates but also the longest average response time.

6. Discussion and Conclusions

The extended MMixIRT model integrates features of IRT models, latent class models, and multilevel models, enabling analysis of strategy category characteristics at the process level while estimating ability values at the individual level. At the process level, latent class models identify problem-solving strategies and explore strategy application patterns. At the individual level, IRT models estimate student abilities. The model' s advantage lies in its simultaneous description of information at both process and individual levels. Strategy analysis at the process level can reveal typical behavioral patterns and thinking characteristics of different groups during problem solving, providing targeted information for improving students' problem-solving abilities. Additionally, the extended MMixIRT model demonstrates good parameter recovery and high classification accuracy, making it applicable for process data analysis.

Applying the extended MMixIRT model to analyze process data from five regions validated the model' s reasonableness and interpretability. First, operational steps could be classified into five strategies reflecting characteristics of students at different ability levels during problem solving. Strategy 4 is the correct problem-solving strategy; higher proportions of students using Strategy 4 in their final step correspond to higher average ability levels. Strategy 2 is the poorest strategy, farthest from the correct path and most time-consuming; higher proportions using Strategy 2 correspond to lower average ability levels. Second, results on strategy application and transitions reflect trial-and-error strategy usage during problem solving, consistent with real-world problem-solving behavior. Among correct-group students, applying 4 or 5 strategies was most common, indicating that students completed tasks through continuous strategy transitions, with few solving problems correctly using only one strategy from the start. The most typical correct-group strategy transition was from Strategy 3 to Strategy 4: students first selected the first three paths overlapping with the correct route, then at the next path, chose Park instead of Lee from Market (see [Figure 3: see original paper]; the latter three paths sum to 16 minutes, longer than the correct answer' s 15 minutes). Students likely discovered a shorter route existed and transitioned from Strategy 3 to Strategy 4, changing from Market to Lee and selecting the correct path. In contrast, error-group students most frequently applied 4 strategies but fewer persisted through all 5 strategies. Finally, analysis of relationships among process variables, strategies, and abilities confirmed the model' s validity for process data analysis. Some process variables significantly correlated with individual problem-solving ability, showing that beyond strategy selection, other process variables (e.g., number of path clicks, number of resets) also correlate with problem-solving ability to varying degrees.

The study also compared results across regions at both process and individual levels. First, regions showed different characteristics on process variables. For example, U.S. students had the lowest problem-solving ability, fewest path clicks, and error groups with significantly larger differences between time spent and correct response time than other regions. Singaporean students had the highest problem-solving ability but also the longest response times. Culturally, Western students focus on individual values, curiosity, and interest, while Confucian-background students emphasize effort as essential for success (Li, 2012) [?]. This item assesses problem-solving planning and execution; persistent trial-and-error can lead to correct answers. Effort (instrumental motivation) promotes problem-solving performance, possibly explaining higher abilities among Singaporean, Shanghai-Chinese, and Hong Kong-Chinese students. Singapore' s superior performance stems from its curriculum design systematically incorporating problem-solving ability into mathematics syllabi, explicitly listing thinking skills and problem-solving strategies (Fan & Zhu, 2007) [?]. Consequently, Singapore' s correct group showed a lower proportion of students applying all 5 strategies, indicating their typical problem-solving characteristic involves longer thinking time, fewer strategies, and correct answers. In contrast, U.S. correct-group students also showed low proportions applying 5 strategies, but their typical characteristic was shorter thinking time without attempting enough strategies to reach correct solutions. Second, error groups in Canada, Hong Kong-China, Shanghai-China, and Singapore more frequently used Strategy 3 in their final step. Since many correct students transitioned from Strategy 3 to Strategy 4, these error-group students would likely convert to correct strategies with more thinking time. These results provide rich information for instruction and targeted guidance.

In summary, process data analysis can provide educational measurement researchers and test developers with more information for item improvement and can be incorporated into test scoring systems. Scoring would no longer rely solely on final responses but would incorporate strategy usage, enriching score meaning.

The extended MMixIRT model is flexible and can be adapted based on item characteristics and research foci. First, latent classes describing student category characteristics can be added at the individual level, or continuous latent variables describing step ability can be added at the process level to explore ability changes during problem solving (Liu, Liu, & Li, 2018) [?]. Second, covariates that reduce measurement error and predict problem-solving ability, such as student motivation, can be included (Fox & Glas, 2003) [?]. Finally, this study involved a single-task scenario; for multiple tasks, the model can be extended to three levels—process, task, and individual—to examine strategy application across different task contexts and ability estimation across multiple tasks.

This study has limitations. First, in strategy transition analysis, using a strategy three or more times was defined as typical strategy usage, losing information about unstable strategy transitions. If such unstable transitions are of interest, they could be incorporated into transition analysis. Second, the analysis treated single paths as units without considering possible path combinations (e.g., when links between different paths are unique, analyzing combined paths might be more reasonable). Future research could consider models for transitions between different path combinations. Additionally, the model' s generalizability to complex problem-solving processes requires further validation. Using MMixIRT requires recoding process data into structures similar to this study, which may not be easily achievable for all tasks.

Main conclusions are:

1. The extended MMixIRT model can analyze strategy usage during problem solving based on behavioral sequences while providing ability estimates at the individual level.
2. The extended MMixIRT model can analyze typical characteristics of strategy usage among students from different regions, providing references for targeted training.

Appendix 1. Class Proportions and Item Correct Response Probabilities in Simulation Study

Three latent classes Five latent classes
(33.33%) (33.33%) (33.33%) (20.00%) (20.00%) (20.00%) (20.00%) (20.00%)

Note: Item correct response probabilities refer to all process steps except the final response status. True values for final response status classification are consistent with each latent class' s characteristics.

Appendix 2. Number of Clicks on Each Path by Class

(Corresponding author: Hongyun Liu, E-mail: hyliu@bnu.edu.cn)

Author Contributions: - Meijuan Li: Data cleaning and analysis; manuscript drafting and final revision - Yue Liu: Manuscript drafting and final revision - Hongyun Liu: Research conceptualization, study design, manuscript drafting and final revision

Note: Figure translations are in progress. See original paper for figures.