

# Predicting Missing Responses with Process Data: A Multiclass Machine Learning Approach

Qiwei (Britt) He, Ph.D.

Data Science and Analytics Program  
AI-Measurement and Data Science (AIMD) Lab  
Georgetown University



# Process Data

- The use of computers as a delivery platform not only enables the development of **innovative item types** but also facilitates the collection of **a broader range of records** in **log files** throughout **human-machine interactions**.
- These granular records, often referred to as *process data*, are typically stored as **an ordered sequence of multi-type, time-stamped events**.
- This rich source of data supports the exploration and identification of informative features from problem-solving processes beyond response data.

PISA 2022

Science Fair Poster  
Question 2 / 2

Use the drawing tools on the right and the text box below to answer the question.

Improve the poster on the right to make it relevant to the theme of Life in Deep Space. Your improvements should be original in the sense that not many people would think of changing the poster in this way. Make sure that the existing poster design is still visible in the final design.

Describe your design in one sentence in the box below.

We recommend that you spend no longer than 5 minutes on this question.

Description

Available Stickers:

PISA 2022

Space Comic  
Question 1 / 2

Refer to the comic strip on the right. Use the text boxes to answer the question.

The text in a comic strip is intended to convey the dialogue between characters. Write a comic strip using the illustrations and text boxes on the right. Your comic strip should be original in the sense that not many people would think of it. Use as many of the available text boxes on the right as needed.

We recommend that you spend no longer than 5 minutes on this question.

Text 1

Text 2

Text 3

Text 4

Text 5

Text 6

4 Experimental Set-up: Fertilizer

Your experiment should be a fair test of whether adding fertilizer to the soil helps the plants produce more peppers and of whether Fertilizer A or Fertilizer B helps the plants produce the most peppers.

Now decide the amount of fertilizer to add to each tank. You can add:

- Fertilizer A only
- Fertilizer B only
- both Fertilizer A and Fertilizer B
- no fertilizer

Click each tank to choose the amount of fertilizer to add.

Tank 1  
Fertilizer (ml)

Tank 2  
Fertilizer (ml)

Tank 3  
Fertilizer (ml)

When you are finished setting up the fertilizer in all three tanks, click

PISA 2022

Spinners  
Question 2 / 3

Refer to "Spinners" on the right. Use the simulator to help you answer the question below. Type your answer to the question.

The theoretical probability that the arrow will stop on any one of the six colours in the spinner shown to the right is  $\frac{1}{6}$ .

As the number of spins increases, how does the percentage of times that the arrow stops on each colour relate to the theoretical probability?

Explain your answer.

Number of Spins	Green	Orange	Purple	Yellow	Blue	Red
50	15 (30.0%)	15 (30.0%)	15 (30.0%)	0	0	0
100	15 (15.0%)	15 (15.0%)	15 (15.0%)	15 (15.0%)	15 (15.0%)	15 (15.0%)
500	71 (14.2%)	81 (16.2%)	64 (12.8%)	105 (21.0%)	90 (18.0%)	89 (17.8%)
1000	69 (6.9%)	170 (17.0%)	178 (17.8%)	156 (15.6%)	160 (16.0%)	164 (16.4%)
5000	809 (16.18%)	809 (16.18%)	809 (16.18%)	804 (16.08%)	828 (16.56%)	805 (16.1%)
5000	833 (16.66%)	833 (16.66%)	809 (16.18%)	813 (16.26%)	853 (17.06%)	842 (16.84%)

PISA 2022

Forested Area  
Practice

You are now going to practise using the spreadsheet below continuing on to the questions.

Use the spreadsheet to complete the following three actions.

- Sort a column.
  - Click on the symbol in Column B, C, or D to sort that column in ascending (low to high) order.
  - Note that all the columns will sort based on the way any one column is sorted.
- Perform a calculation.
  - Select a column from the first drop-down menu located below the spreadsheet.
  - Next, select an operation from the middle drop-down menu.
  - Then select a column from the last drop-down menu.
  - Click on "Run".
  - The results will display in the first available empty column.
- Display the mean (average) of a column.
  - Select a column from the drop-down menu next to "Mean," located below the spreadsheet.
  - Click on "Run".
  - The result will display in the cell below that column.

Country	2005	2010	2015
Algeria	5.86	5.81	5.82
Argentina	11.77	11.74	11.77
Colombia	54.26	52.85	52.73
Germany	32.66	32.73	32.76
Greece	29.31	29.28	29.45
India	22.77	23.47	23.77
Kazakhstan	1.24	1.23	1.23
Lithuania	13.34	13.38	13.42
Paraguay	64.33	63.21	62.11
Peru	59.01	58.45	57.79
Portugal	36.52	35.89	35.25
Singapore	69.65	64.51	62.97
South Korea	64.42	64.08	63.69
Thailand	31.51	31.81	32.1
United States	33.26	33.7	33.85

Calculate

Column C  Multiply  Column B  Run

Mean  Column B  Run Clear All

PISA 2022 Creative Thinking Example Item

TIMSS 2019 Scientific Problem-Solving and Inquiry Example Item

PISA 2022 Mathematics Example Item

# Examples from Different Subjects

# PISA Dataset

90 countries × 5000 students × 100 items × 10 transitions × 20 (actions and time point per item)  
=9,000 million

400 times

45 countries × 5000 students × 100 items  
=22.5 million



PISA 2012



PISA 2022

# Why Does Response Process Matter?

Task  
construction

Investigating  
if examinees  
interact with  
tasks as  
intended

Invariance  
explorations

Defining  
invariance  
between  
group not  
only by item  
difficulty but  
also problem-  
solving  
process and  
strategies

Richer  
description

Investigating  
if examinees  
could solve a  
given task  
and how  
efficiently  
they did

Refining  
theories

Providing  
evidence for  
cognitive  
process that  
could not be  
captured  
from paper-  
pencil based  
assessments

Tailored  
interventions

Identifying  
subskills or  
meta  
competencies  
that  
examinees  
are lacking



## Prepare for Digital Age

- Changes in society, environment, and technology are **shifting the emphasis of education** from equipping students with routine skills to empowering them to confront and overcome complex challenges in a **digital world**.
- **Computational thinking** and **adaptive problem-solving skills** in digital tasks have been recognized across countries as among the most important skills in the 21st century.

# Case study: International Computer and Information Literacy Study (ICILS)



- The ICILS 2018 organized by the IEA extends the evaluation of students' **computer and information literacy (CIL) skills** and introduces a novel assessment of students' **computational thinking (CT) skills**.
- CT is defined as the ability to recognize, analyze, and describe real-world problems so their solutions can be operationalized within **programming tasks** (Fraillon et al., 2019).
- Students in the 8<sup>th</sup> year of schooling
- 9 countries and regions

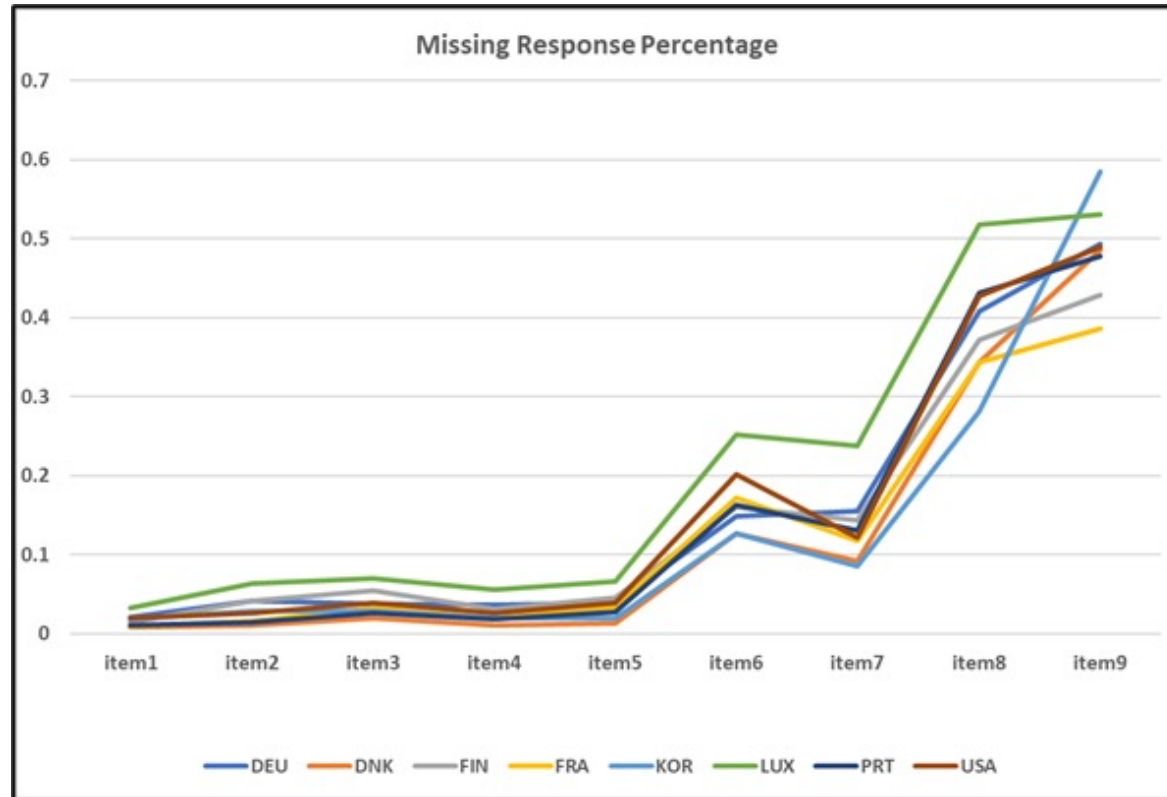
# Instrument: Farm Drone CT Module

- **Correctness:** Degree to which the farm drone (under programming code control) performs as required and presence or absence of any unrequired actions.
- **Efficiency:** Programming code is compared with the number of codes in minimum to achieve a fully correct solution, unnecessary longer code sequences corresponding to lower scores.



Task No.	Task type	Target layout	Number of Targets	Average Time (Seconds)	Average Number Command	Average Reset	Average Correctness	Average Nonresponse Percentage
Task 1	Warm up	N/A	N/A	76.9	3.5	2.0	95.1%	1.8%
Task 2	Creation	single	1	40.9	4.0	0.9	88.3%	3.3%
Task 3	Debugging	single row	4	119.0	8.1	2.6	84.8%	4.0%
Task 4	Creation	single row	4	99.4	10.0	1.5	85.6%	3.0%
Task 5	Creation	double row	8	236.1	16.6	4.1	76.7%	3.9%
Task 6	Debugging	single row	6	127.9	8.9	2.5	62.3%	17.8%
Task 7	Creation	double row	12	195.6	16.4	3.6	50.9%	14.4%
Task 8	Debugging	quadruple row	16	120.4	16.7	3.9	26.7%	40.8%
Task 9	Configuration	quadruple row	16	78.9	NA	3.3	35.5%	49.0%

# Missing Responses



Among the **31,344 students**, 11,468 (36.6%) students gave full responses to all the nine tasks, while **19,876 (63.4%)** students had **at least one missing response** in the 9 tasks.

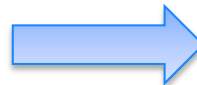
# Research Questions

- Where the students were blocked to give responses? Are there any meaningful patterns for the missing values?



Clustering students' missing response patterns  
*(with 19,876 students who had at least one missing response in the 9 CT tasks)*

- Could we use process data to help recover missing values to improve the CT latent trait estimation?



Predicting missing responses with process data  
*(with 11,468 students who gave full responses throughout the 9 CT tasks)*

# Sequence Mining

- Sequential pattern mining is a topic of data mining concerned with finding statistically relevant patterns between data examples where the values are **delivered in a sequence**.
- Sequence mining techniques include building efficient databases and indexes for **sequence information**, **extracting the frequently occurring patterns**, **comparing sequences for similarity**, and **recovering missing sequence members**.

Dong & Pei (2007)

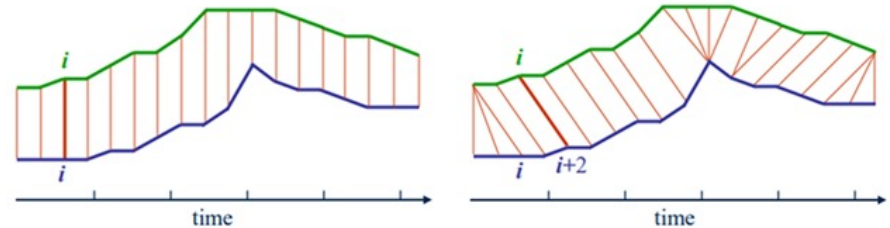
# Sub-Study 1: Clustering Missing Response Patterns with Dynamic Time Warping (DTW) Method

- Dynamic time warping (Sakoe & Chiba, 1978) is a distance measure that searches the optimal warping path between two series.

- Given sequences

$$X = \{x_1, x_2, \dots, x_n\} \text{ and}$$

$Y = \{y_1, y_2, \dots, y_m\}$  with the same or different lengths, a warping path  $W$  is an alignment between  $X$  and  $Y$ , involving one-to-many mappings for each pair of elements.



Any distance (Euclidean, Manhattan, ...) which aligns the  $i$ -th point on one time series with the  $i$ -th point on the other will produce a **poor similarity score**.

A non-linear (elastic) alignment produces a **more intuitive similarity measure**, allowing similar shapes to match even if they are out of phase in the time axis.

He, Q., Borgonovi, F., Suárez-Álvarez, J. (2023). Clustering Sequential Navigation Patterns in Multiple-Source Reading Tasks with Dynamic Time Warping Method. *Journal of Computer-Assisted Learning*, 39(3), 719-736.

# Dynamic Time Warping Algorithm

- The initial step of DTW algorithm is defined as:

$$DTW(i, j) = \begin{cases} \infty & \text{if } (i = 0 \text{ or } j = 0) \text{ and } i \neq j \\ 0 & \text{if } i = j = 0 \end{cases}$$

- The recursive function of DTW is defined as

$$DTW(i, j) = \min \begin{cases} DTW(i - 1, j) + w_h C(i, j) \\ DTW(i, j - 1) + w_v C(i, j) \\ DTW(i - 1, j - 1) + w_d C(i, j) \end{cases}$$

where  $(w_h, w_v, w_d)$  are weights for the horizontal, vertical and diagonal directions, respectively.  $DTW(i, j)$  denotes the distance or cost between two subsequences  $\{x_1, x_2, \dots, x_i\}$  and  $\{y_1, y_2, \dots, y_j\}$ , and  $DTW(N, M)$  indicates the total cost of the optimal warping path.

	5	10	6	3	1	2	4	3
	4	6	3	1	0	1	3	3
A	3	3	1	0	1	1	2	4
	2	1	0	1	3	4	4	7
	1	0	1	3	6	8	9	13
		1	2	3	4	3	2	5
		B						

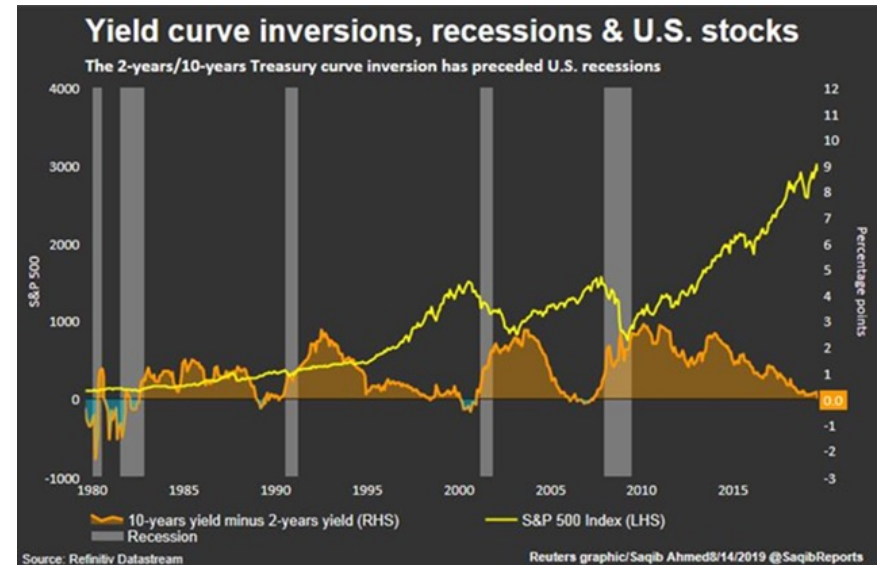
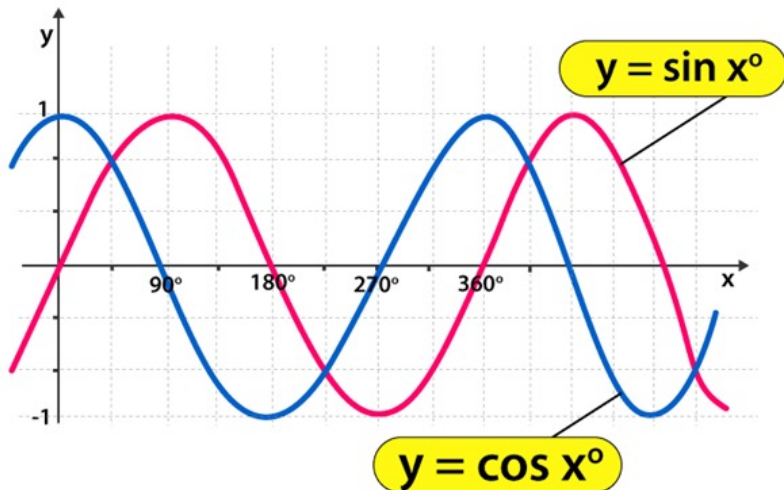
$A_i = \{1, 2, 3, 4, 5\}$

$B_j = \{1, 2, 3, 4, 3, 2, 5\}$

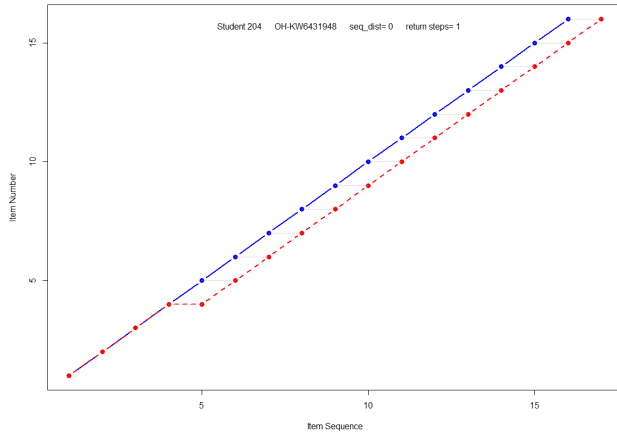
$$\begin{aligned} dtw(i, j) &= |A_i - B_j| + \min(D[i - 1, j - 1], D[i - 1, j], D[i, j - 1]) \\ &= |5 - 2| + 3 \end{aligned}$$

Recode categorical variables to numeric variables (item response score=[0,4], missing response=9)

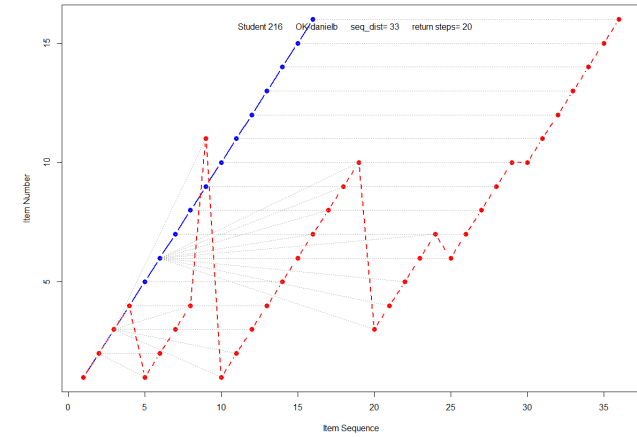
# Reasons to Use DTW



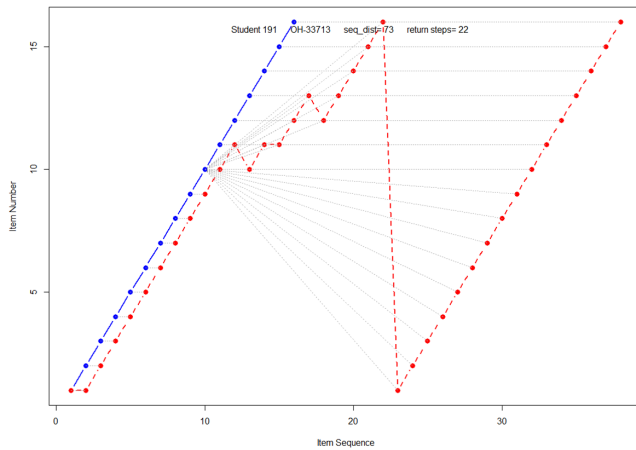
# Use of DTW: Visit-Revisit (Navigation) Sequence Distance



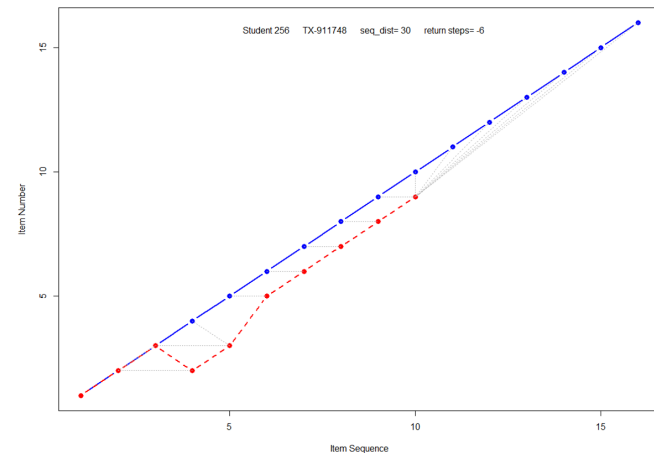
**normal pattern**



**multiple back-forth revisit pattern**

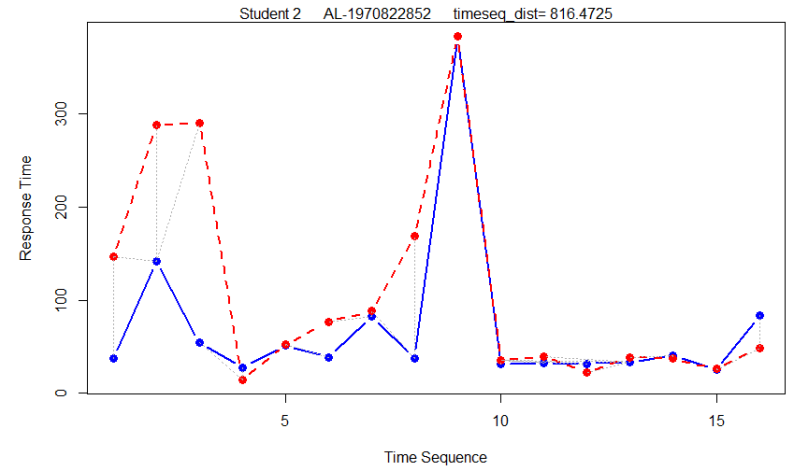
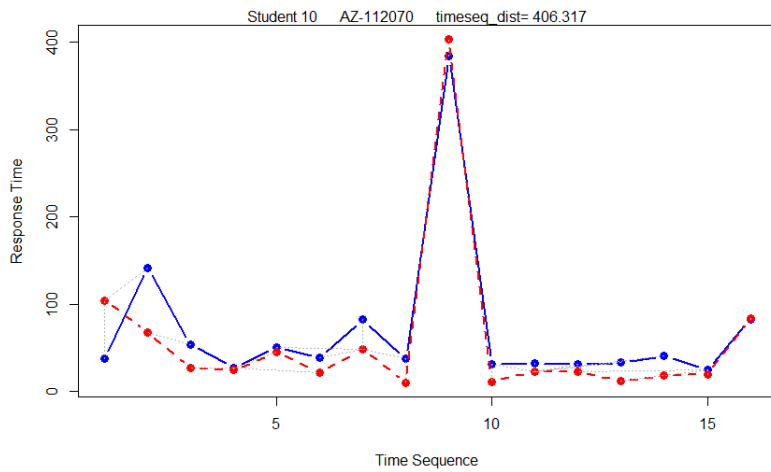


**double review pattern**



**quickly skip pattern**

# Use of DTW: Time Sequence Distance



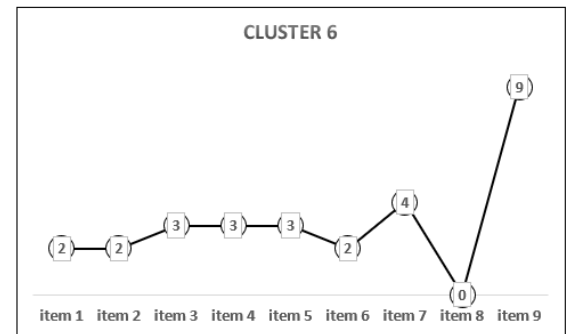
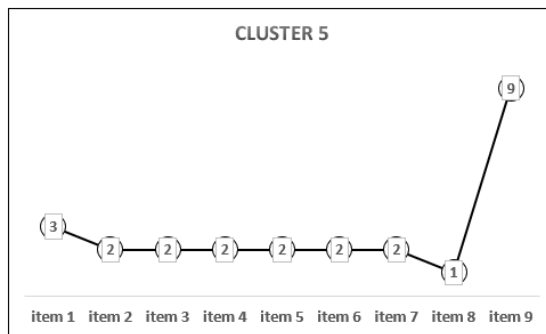
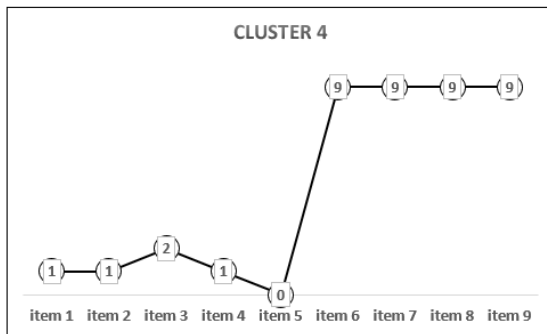
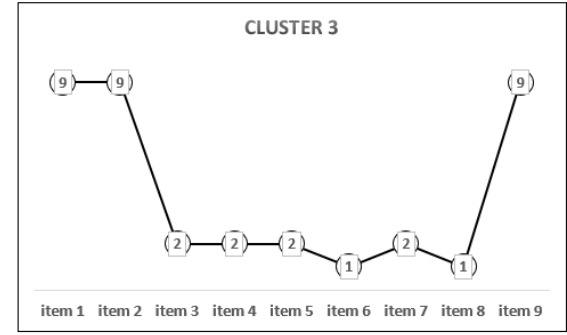
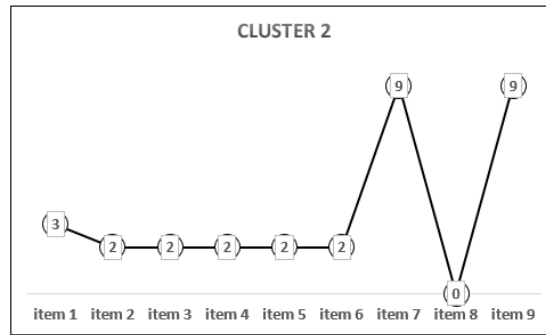
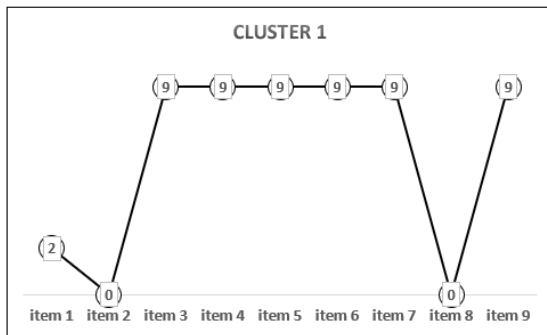
# Process Data Support CT Skills Measurement (He & Gonzalez, 2023)



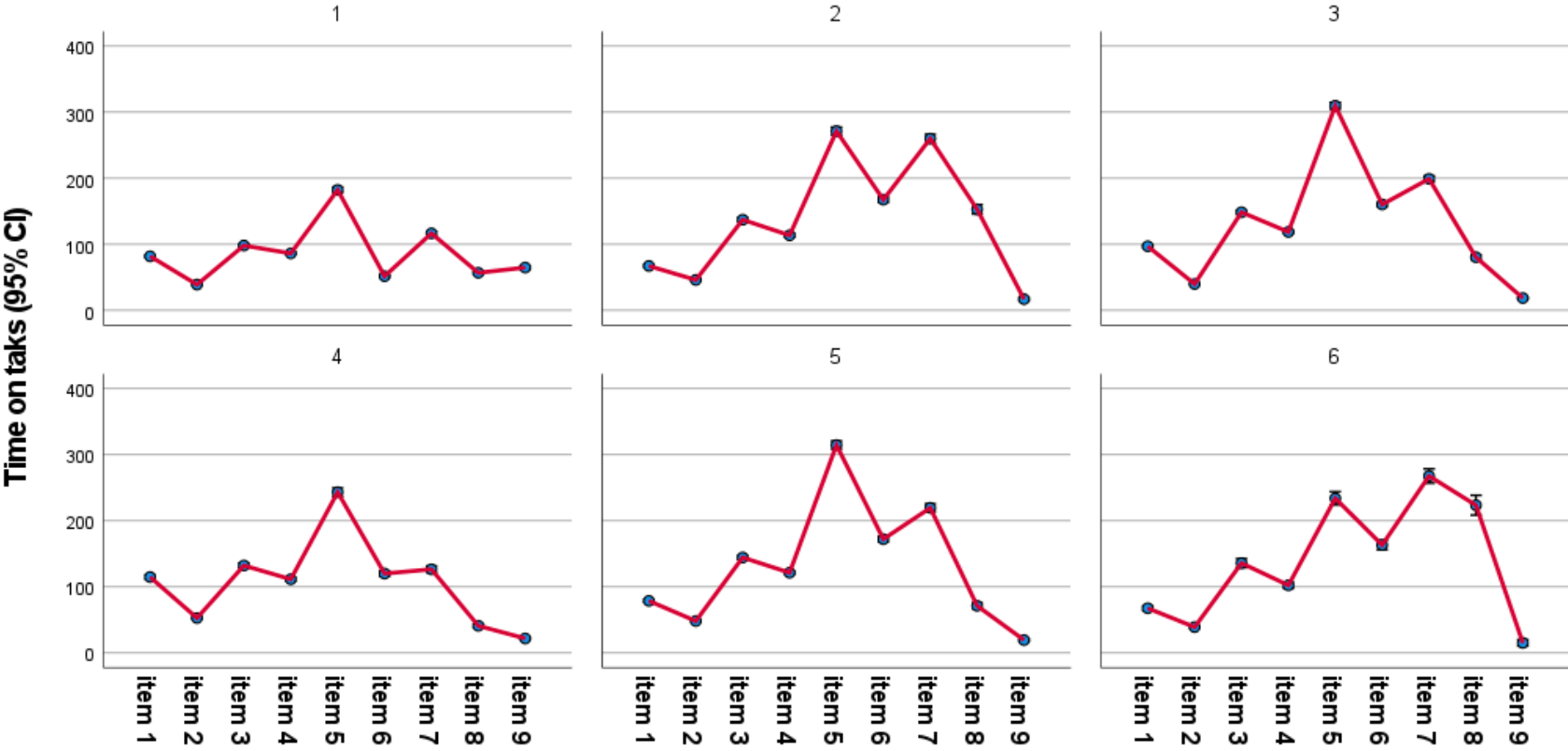
A screenshot of a programming environment titled 'Work Space:6'. On the left is a 10x10 grid representing a field with green grass and brown dirt. A small drone icon is on the grid. In the center is a code editor with the following blocks: 'when run' (orange), 'repeat 2 times' (pink) containing 'move forward' (teal), 'turn right' (teal), 'repeat 4 times' (pink) containing 'move forward' (teal) and 'drop water' (teal). On the right is a 'Time Remaining' section showing '00 mins' and a vertical bar with colored segments (green, white, blue). At the bottom left, a 'Code Blocks' panel contains 'move forward', 'turn left', 'drop water', and 'repeat 3 times' blocks. At the bottom, there are instructions: 'Use the 'repeat do' code block to make the drone drop water on the four dirt tiles with seeds. The drone should not drop water on any grass tiles.' and 'Use as few code blocks as possible to complete the task. Click [play icon] to see the results. Click [right arrow icon] when you are ready to continue.' A green arrow button and a blue search button are at the bottom right.

- 19,876 students (over 2/3 of the total sample) who had at least one missing response throughout the Farm Drone module with the aim to **extract the nonresponse sequential pattern** and **pinpoint the potential reason for the missing responses**.
- Recode score sequence e.g., (1, 2, 3, 0, 9, 0, 1, 3, 9)
- Dynamic Time Warping (DTW)
- Sequence clustering based on pairwise sequence distance matrix, the maximum value of Silhouette index to determine the optimal number of clusters.

# Nonresponse Patterns



# Time Allocation Patterns



## Sub-Study 2: Predicting Missing Responses with Process Data via Machine Learning Methods

- This study used 11,468 students who gave full responses throughout the 9 CT tasks, with the aim to get “gold standard” to evaluate the prediction performance
- Two machine learning methods: Random Forest (RF) and Support Vector Machine (SVM)
- To predict the missing response by each item using the rest 8 items’ process data and response data both. 80% training and 20% test, with cross validation.
- Multiple class prediction: each item has multiple but unequal classes (partial credits may differ across tasks), min = 0, max = 4

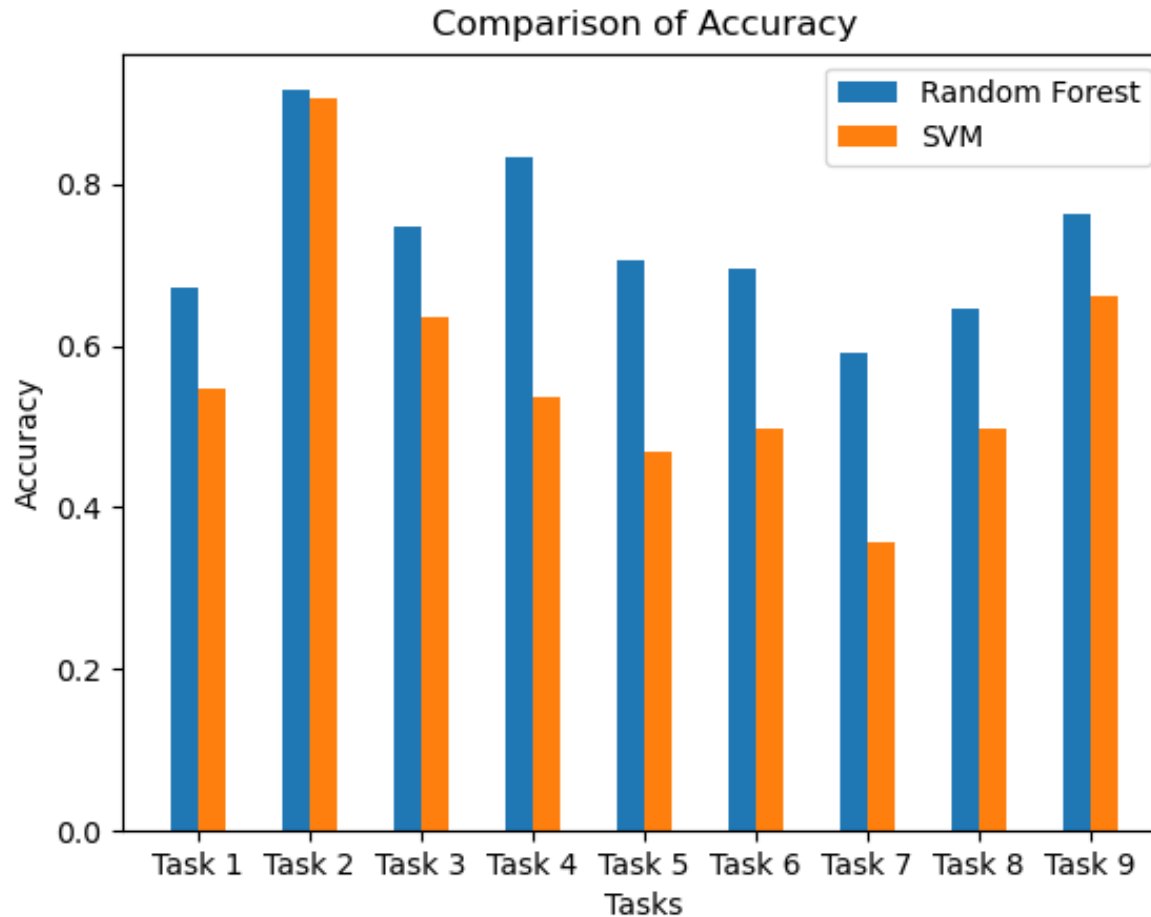
Variables	Number of variables	Tasks Involved	Description
Algorithms	6	3,4,5,6,7,8	The programming coding strategy that is chosen by student, either by straightforward move codes or by nested codes. There are 6 levels predefined in this variable. Only one algorithm strategy could be labeled per student per item. Level 1: move with turn/drop Level 2: nested move in a repeat Level 3: nested move with drop in a repeat Level 4: nested if do in a repeat Level 5: nested repeat in a repeat level 6: nested if do in a repeat nested in another repeat
Commands	8	1,2,3,4,5,6,7,8	The number of command lines recorded in the item (the final commands shown on screen only. The adding or removing commands during the problem-solving process are not counted.
Irrelevant Targets	6	3,4,5,6,7,8	A binary indicator to record “one or more irrelevant target” (label 0) and “No irrelevant targets” (label 1)
Remaining time	1		The time remaining in the Farm Drone module after submitted responses to all 9 items. A total 25 minutes is set for the whole module.
Reset	9	1,2,3,4,5,6,7,8,9	Number of clicks on Reset button to restart the Farm Drone simulation. (Only the Reset button in the simulation space is counted.)
Task Time	9	1,2,3,4,5,6,7,8,9	Total time spent on an individual task. Students are allowed to get back and forth to different task within one module. The task time is a sum of time spent in every visit.

## Process Variables Captured in ICILS CT module

Process data were collapsed during the data collection process.

Only aggregate level process data are available, but not sequential fine-grained data.

# Prediction Results (1) – Macro-Level Accuracy



# Prediction Results (2) – Micro-Level Accuracy

		Precision	Recall	F1
Task 1	Class 0	0.960	0.192	0.305
	Class 1	0.648	0.330	0.437
	Class 2	0.653	0.610	0.631
	Class 3	0.684	0.802	0.738
Task 2	Class 0	0.731	0.438	0.548
	Class 1	0.964	0.121	0.215
	Class 2	0.920	0.998	0.957
Task 3	Class 0	0.785	0.594	0.676
	Class 1	0.766	0.277	0.405
	Class 2	0.785	0.828	0.806
	Class 3	0.657	0.734	0.693
Task 4	Class 0	0.844	0.629	0.720
	Class 1	0.762	0.195	0.309
	Class 2	0.800	0.874	0.835
	Class 3	0.869	0.894	0.881
Task 5	Class 0	0.831	0.466	0.597
	Class 1	0.624	0.642	0.633
	Class 2	0.658	0.584	0.619
	Class 3	0.760	0.875	0.814
Task 6	Class 0	0.710	0.436	0.539
	Class 1	0.587	0.571	0.578
	Class 2	0.750	0.805	0.776
	Class 3	0.707	0.843	0.769
Task 7	Class 0	0.613	0.780	0.687
	Class 1	0.583	0.161	0.252
	Class 2	0.442	0.353	0.393
	Class 3	0.513	0.676	0.583
	Class 4	0.721	0.711	0.716
Task 8	Class 0	0.668	0.816	0.734
	Class 1	1.000	0.124	0.219
	Class 2	0.578	0.511	0.542
	Class 3	0.650	0.598	0.623
	Class 4	0.767	0.892	0.825
Task 9	Class 0	0.767	0.892	0.825
	Class 1	0.969	0.117	0.209
	Class 2	0.746	0.699	0.721

- In the first 7 tasks, the high performance classes are more likely to be correctly predicted, which contribute more to the macro-level accuracy.
- In the last 2 items, the lowest performance classes are more likely to be correctly predicted, which contribute more to the macro-level accuracy.
- These results show that the process information may provide more predictable information to the high performance groups in most tasks, while approaching to the end of the unit, the low performance groups have more dominant information.

# Prediction Results (3) – Predictable Variables by Task

## Top 5 Informative Variables in Missing Value Predictions with Random Forest

	Task1	Task2	Task3	Task4	Task 5	Task 6	Task 7	Task 8	Task 9
Feature 1	T2F02T	T2F01T	T2F04AA	T2F05AA	T2F04AA	T2F07AA	T2F06AA	T2F09T	T2F05AA
Feature 2	T2F05T	T2F05T	TF04TEC	TF05TEC	TF04TEC	TF07EC	TF06TEC	T2F07T	TF05TEC
Feature 3	T2F08T	TF01E	T2F04T	T2F03AA	T2F04ZA	T2F07XA	T2F08T	T2F05T	T2F03AA
Feature 4	T2F06T	T2F04T	T2F05T	TF03TEC	T2F07T	T2F07T	T2F05T	T2F04T	TF03TEC
Feature 5	T2F04T	T2F09T	T2F07T	T2F03ZA	T2F04T	T2F08T	T2F06T	T2F	T2F03ZA

- Information from response time (T) and command actions (AA) dominate in the top feature, superior than scores of other items (EC).
- Adjacent items provide more information than other items in missing value prediction.

# Conclusions and Future Studies

- Process data can provide more information in missing value prediction, thus, worthwhile to include them for a better estimation.
- Process variables related to response time and action lengths can superior than scores of other items in missing value prediction.
- Process data in adjacent items to the missing items are more likely to provide higher information than other items.

# Final thoughts for process data studies in the future

- Process data (**personalized behavioral data**) may add more weights in assessment model in the future to compensate the risk of AI manipulation. Behavioral data could be more informative than self-reported scale.
- One-time assessment is not the target of education, but tracking the students' progress and providing timely and helpful supports. Process data could be given a broader scope for **dynamic and longitudinal study**. Sequence methods could also be used in educational policy tracking.
- Post-assessment process data analysis could be progressed to dynamic process data analysis throughout the assessment, bringing **new meaning for adaptive testing**, not only from responses but also from interactions.

# Reference

- Han, Z., He, Q., & von Davier, M. (2019). Predictive Feature Generation and Selection Using Process Data from PISA Interactive Problem-Solving Items: An Application of Random Forests. *Frontiers in Psychology, 10*: 1421.
- Hao, J., Shu, Z., & von Davier, A. (2015). Analyzing process data from game/scenario-based tasks: an edit distance approach. *Journal of Educational Data Mining, 7*(1), 33-50.
- He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Identifying generalized behavioral patterns with sequence mining. *Computers and Education, 166*, 104170.
- He, Q., Borgonovi, F., Suárez-Álvarez, J. (2022). Clustering Sequential Navigation Patterns in Multiple-Source Reading Tasks with Dynamic Time Warping Method. *Journal of Computer-Assisted Learning*. DOI: 10.1111/jcal.12748
- He, Q., & von Davier, M. (2016). Analyzing Process Data from Problem-Solving Items with N-Grams: Insights from a Computer-Based Large-Scale Assessment. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.) *Handbook of Research on Technology Tools for Real-World Skill Development* (pp. 749-776). Hershey, PA: Information Science Reference.
- Heldt, M., Masek, C., Drossel, K., & Eickelmann, B. (2020). The relationship between differences in students' computer and information literacy and response times: an analysis of IEA-ICILS data. *Large-scale Assessments in Education, 8*(1), 1-20.
- Tang, X. (2023). A Latent Hidden Markov Model for Process Data. *Psychometrika, 1*-36.
- Tang, X., Wang, Z., He, Q., Liu, J., & Ying, Z. (2020). Latent feature extraction for process data via multidimensional scaling. *Psychometrika, 85*(2), 378-397.
- Tang, X., Wang, Z., Liu, J., & Ying, Z. (2021). An exploratory analysis of the latent structure of process data via action sequence autoencoders. *British Journal of Mathematical and Statistical Psychology, 74*(1), 1-33.
- Ulitzsch, E., He, Q., Pohl, S. (2022). Using sequence mining techniques for understanding incorrect behavioral patterns on interactive tasks. *Journal of Educational and Behavioral Statistics, 47*(1), 3-35.
- Ulitzsch, E., He, Q., Ulitzsch, V., Nichterlein, A., Molter, H., Niedermeier, R., Pohl, S. (2021). Combining clickstream analyses and graph-modeled data clustering for identifying common response process using time-stamped action sequence. *Psychometrika, 86*, 190-214.
- Xiao, Y., He, Q., Veldkamp, B., & Liu, H. (2021). Exploring Latent States of Problem-Solving Competence Using Hidden Markov Modeling on Process Data. *Journal of Computer-Assisted Learning, 37*(5), 1232-1247.
- Zhang, S., Tang, X., He, Q., Liu, J., & Ying, Z. (2024). External Correlates of Adult Digital Problem-Solving Process: An Empirical Analysis of PIAAC PSTRE Action Sequences. *Zeitschrift für Psychologie*.



Thank you very much!

Qiwei (Britt) He, PhD

[qiwei.he@georgetown.edu](mailto:qiwei.he@georgetown.edu)



GEORGETOWN UNIVERSITY