UNIVERSITY OF ALBERTA

Systematic review of the use of process data in the computer-based assessment

Surina He, Ying Cui

Measurement, Evaluation, and Data Science

University of Alberta, Canada

Content

- Background
- Methods
- Results
- Conclusion

Background

- Process data (i.e., log data) typically refers to data stored in log files generated by respondents' interactions with a testing platform. Information from process data is valuable for educational research and practice because it could provide fine-grained response processes and timestamps when students interact with computer-based assessment system.
- Research about the use of process data in the educational assessment increasingly developed recent years. But there is few systematic review about it.
- Therefore, the current systematic review aims to examine following research questions:
  - What types of process data have been used to examine test-takers' response process?
  - What kind of scientific topics have been investigated using process data and what are their findings?
  - What are the typical procedures for pre-processing process data?
  - What are the common approaches for analyzing and modelling process data?

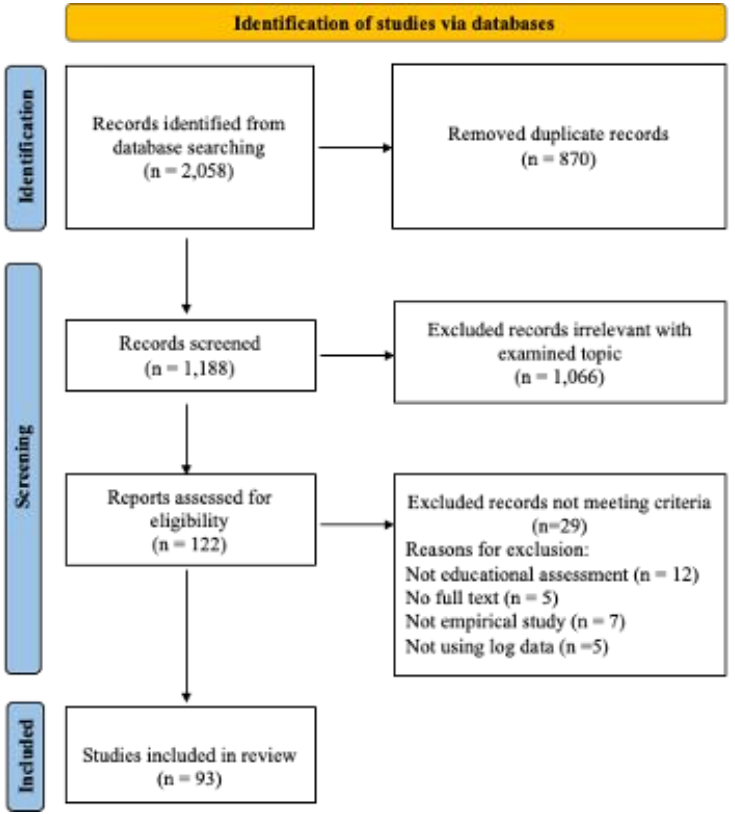
Methods

- Followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines
- Search terms →
  - Process data
  - Educational measurement

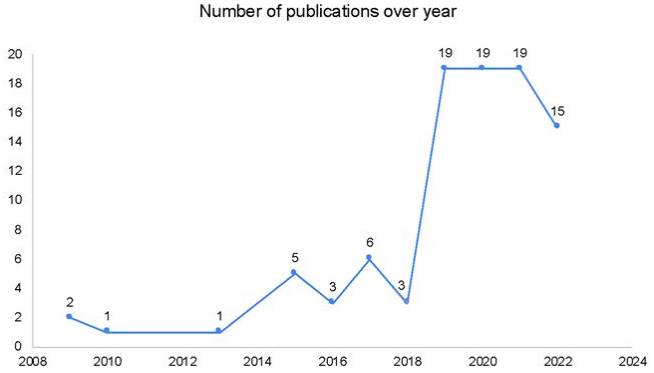
Database	Search strategy	Number of records
Web of Science - All Databases	1. "process data" or "log data" or "trace data" or "event data" or "event log*" or "sequential data" OR "response process*" AND 2. PISA or "Program* for International Student Assessment" or PIAAC or "Program* for the International Assessment of Adult Competencies" or TIMSS or "Trend* in International Mathematics and Science Study" or PIRAL or "Progress in International Reading Literacy Study" or NAEP or "National Assessment of Educational Progress" or "summative assessment*" or "formative assessment*" or ((educat* OR student* OR learner*) NEAR/3 (assess* OR evaluat* OR measur*)) or (large scale NEAR/3 (assess* OR measur*))	804
Scopus	1. "process data" or "log data" or "trace data" or "event data" or "event log*" or "sequential data" OR "response process*" AND 2. PISA or "Program* for International Student Assessment" or PIAAC or "Program* for the International Assessment of Adult Competencies" or TIMSS or "Trend* in International Mathematics and Science Study" or PIRAL or "Progress in International Reading Literacy Study" or NAEP or "National Assessment of Educational Progress" or "summative assessment*" or "formative assessment*" or ((educat* OR student* OR learner*) W/3 (assess* OR evaluat* OR measur*)) or (large scale W/3 (assess* OR measur*))	655
EBSCO host 1) ERIC 2) Academic search complete 3) Education index retrospective 4) Education research complete	1. "process data" or "log data" or "trace data" or "event data" or "event log*" or "sequential data" OR "response process*" AND 2. PISA or "Program* for International Student Assessment" or PIAAC or "Program* for the International Assessment of Adult Competencies" or TIMSS or "Trend* in International Mathematics and Science Study" or PIRAL or "Progress in International Reading Literacy Study" or NAEP or "National Assessment of Educational Progress" or "summative assessment*" or "formative assessment*" or ((educat* OR student* OR learner*) N3 (assess* OR evaluat* OR measur*)) or (large scale N3 (assess* OR measur*))	430
Proquest - Education database	1. "process data" or "log data" or "trace data" or "event data" or "event log*" or "sequential data" OR "response process*" AND 2. PISA or "Program* for International Student Assessment" or PIAAC or "Program* for the International Assessment of Adult Competencies" or TIMSS or "Trend* in International Mathematics and Science Study" or PIRAL or "Progress in International Reading Literacy Study" or NAEP or "National Assessment of Educational Progress" or or "summative assessment*" or "formative assessment*" or ((educat* OR student* OR learner*) NEAR/3 (assess* OR evaluat* OR measur*)) or (large scale NEAR/3 (assess* OR measur*))	169

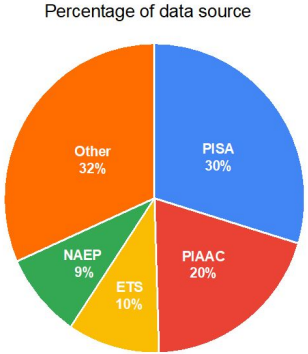
Methods

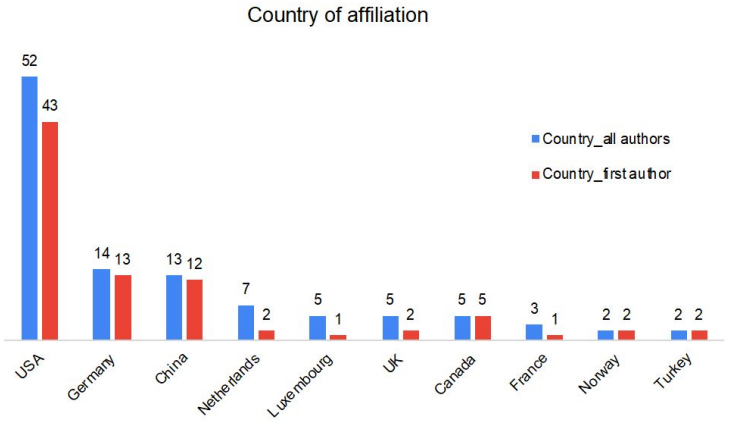
- Peer-reviewed journal articles, books, book chapters, dissertations, and conference proceedings
- Time: until November 2022
- Used Covidence to manage references
- Number of publications in each stage →

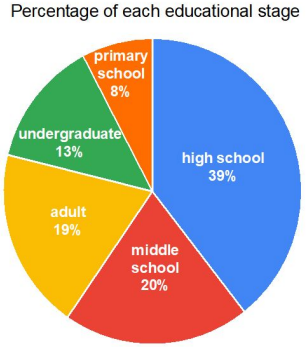


Results: General trend









Results: Type of data

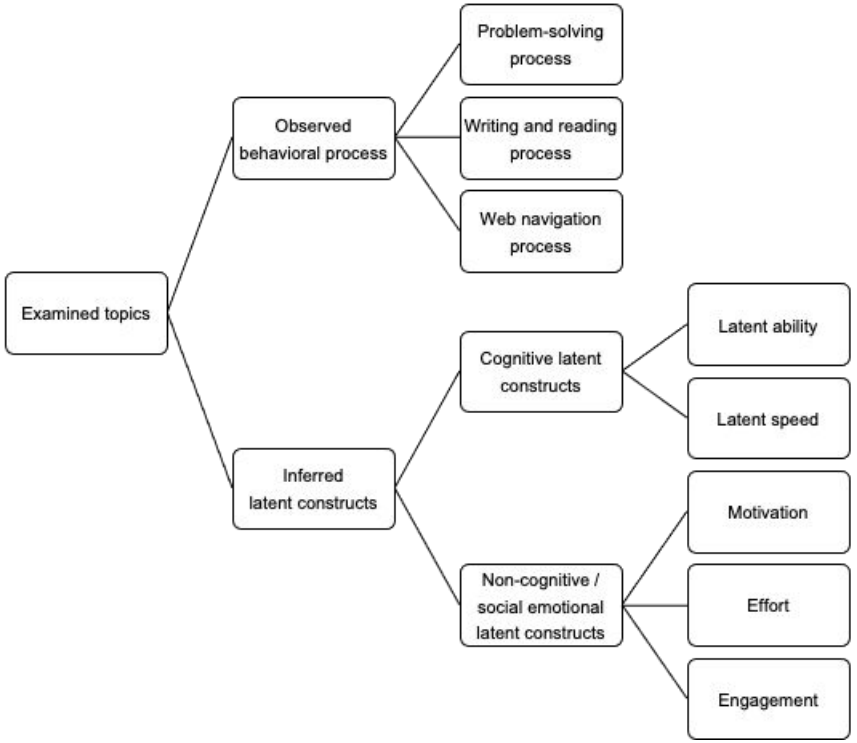
- Type of process data used:
  - Response time 41%
  - Action frequency 30%
  - Action sequence 29%

  (They can be used jointly)
- Potential indicators for each data type →

Data type	Indicators	References
Response time	Total/average time on each item	Bolsinova & Tijmstra, (2019); Goldhammer et al., (2017); Kroehne et al., (2019)
	Total/average time on task/subtask	Goldhammer et al., (2017); Lundgren & Eklöf, (2020); Stadler et al., (2020)
	Total time of each student	Soland, (2019)
	Total/average time for each action	Cui et al., (2019); Ulitzsch et al., (2021)
	Total time for student viewing the questions until Submitting correct/incorrect answers	Papamitsiou & Economides, (2015)
	Time spend on each page	Goldhammer et al., (2017); Yang et al., (2021)
	Duration time from start to specific action	Jiang et al., (2021); Xu et al., (2020)
	Time since last successful/unsuccessfull attempt	Rowe et al., (2021)
	Total/median duration time between two actions	Eichmann et al., (2019); Zhu et al., (2019)
	Variance of duration time between two actions	Eichmann et al., (2019)
	The ratio of time spent on the relevant page to the Total time on task	Goldhammer et al., (2017)
	Timestamp/timepoint	Chen et al., (2019); Liu et al., (2018); Teig et al., (2020)
	Time sequence	Chen, 2020; Fang & Ying, (2020)
Action frequency	Total/average number of specific action (e.g., doub-leclick, re-view, using help button, etc.)	Ivanova et al., (2020); Tate & Warschauer, (2019)
	Total/average number of all action	Eichmann et al., (2020)
	Total/average number of changing answer	Jiang et al., (2021); Papamitsiou et al., (2015)
	Total number of attempting item/task	Harindranathan & Folkestad, (2019); Zehner et al., (2021)
	Total number of action for clicking specific glossary	Wolf et al., (2022)
	Ratio of the action x frequency divided by the action y frequency	Guo, (2022)
	Number of action per minute	Lundgren & Eklöf, (2020)

Results: Examined topics

- Examined topics →
- Under each topic, prior studies explored different research questions by clustering, classifying, or making predictions.



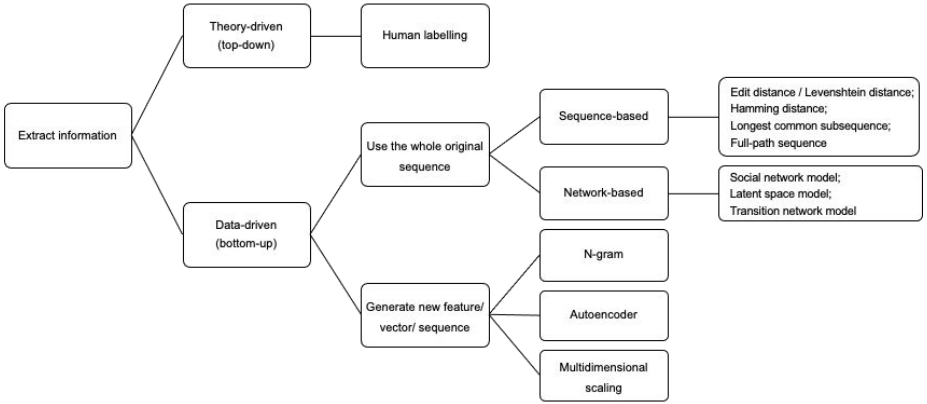
Results: Pre-processing and analyzing

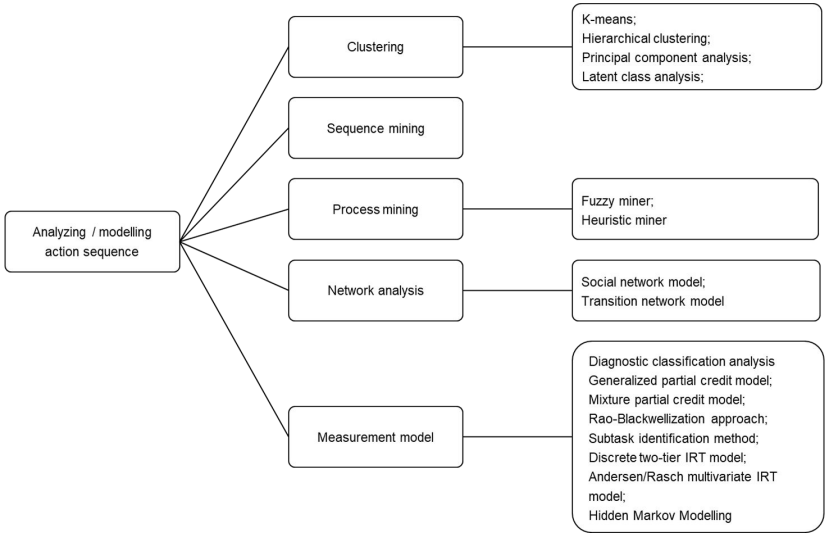
- Response time
  - Pre-processing: log-transformation
  - Analyzing:
    1. Statistical and machine learning analysis
    2. Measurement model:
       Differential response time model;
       IRT-latent regression model;
       Multidimensional IRT model;
       Generalized linear mixed model;
       Social combination IRT model;
       Bivariate generalized linear IRT model

- Action frequency
  - Pre-processing: z-standardization, generating new variables
    (effort = α (average idle time on task / average total time to answer) + β (average total number of times changing answers/average total check views) + γ)
  - Analyzing:
    1. Statistical and machine learning analysis
    2. Clustering
    3. Measurement model:
       Differential item action;
       Explanatory item response modeling;
       Finite mixture model

Results: Pre-processing and analyzing

- Action sequence data is more complex than response time and action frequency.
- So, the important step in the pre-processing procedure is to extract information.





Conclusion

- The current study detaily examined the use of process data in the educational assessment. We organized the types of process data, examined topics, pre-processing procedures, and data analyzing / modelling approaches in the previous empirical studies.
- This review can be helpful to anyone who has no idea about process data but would like to know the current development in this research area. In addition, the classification of three types of process data may be beneficial for researchers develop more appropriate measurement models.
- The limitations of this study includes:
  - Not included specific terms such as “response time”, “number of action” into search terms.
  - Classification for data type, pre-processing and analyzing approaches are subjective.

Thank you for your listening !

For questions and comments, please contact:

Surina He surina1@ualberta.ca

Ying Cui yc@ualberta.ca