ISLS Annual Meeting 2025
June 10-13, 2025

Educating for world-making: Envisioning and enacting sustainable solutions to global crises

18th International Conference on Computer-Supported Collaborative Learning (CSCL)

CSCL Proceedings

Edited by: Jun Oshima, Bodong Chen, Freydis Vogel & Sanna Järvelä

ISLS ANNUAL MEETING 2025 HELSINKI FINLAND
June 10–13, 2025
Helsinki, Finland
Supported by Helsinki
UNIVERSITY OF HELSINKI
FACULTY OF EDUCATIONAL SCIENCES

International Society of the Learning Sciences

Using Markov Chains to Understand Clinical Decision-Making Process

Hanxiang Du, Western Washington University, duh2@wwu.edu
Vitaliy Popov, University of Michigan, vipopov@umich.edu

Abstract: Clinical reasoning is a critical yet complex cognitive process of diagnostic and therapeutic decision-making in medical practice that has long challenged precise understanding and assessment. Sequential analysis can be used to uncover patterns and trends in clinical practices, contributing to improved training and ultimately leading to better patient care outcomes. In this study, 21 board-certified anesthesiologists participated in a simulated-based learning scenario requiring them to promptly recognize patient's condition and initiate appropriate treatment. They were assigned into either the low-performing or high-performing group based on their performance. We utilized Markov Chain Transition Matrix, a robust statistical model for sequential data, to analyze participants' practices using team reflection behavioral observation system and identified statistically significant differences between their transition matrices. The high-performing group had a much higher transition probability from evaluating information to implementation and from planning to planning. The implications are then discussed.

Introduction

Clinical reasoning is the systematic process by which healthcare professionals gather, interpret, and analyze patient information to make informed diagnostic and treatment decisions, integrating their own (biomedical and clinical) knowledge with initial patient information (Gruppen, 2016; Schmidt & Boshuizen, 1993; Richters et al., 2024; Smith et al., 2008). These complex skills are challenging to teach, master, and assess (Gruppen, 2016). Traditional assessment methods like written assessments (e.g., multiple-choice questions, script concordance tests, and post-encounter notes), performance-based assessments (e.g., direct observation, standardized patient encounters, and simulations), and process-focused assessments (think-aloud protocols, chart-stimulated recall, concept mapping) have been fundamental in measuring different aspects of a clinician's diagnostic and therapeutic decision-making abilities (Daniel et al., 2019). These methods, commonly linked with inferential statistical techniques like ANOVA and t-tests, have offered important insights into group differences regarding learners' behaviors and self-reported data. However, these approaches frequently fail to adequately capture the dynamic aspects of clinical decision-making, which involve nuanced differences in the sequence of actions taken by clinicians in a certain scenario (Banning, 2008; Smith et al., 2008).

During clinical diagnostic and treatment reasoning process, diverse states of clinical management occur in sequences and potentially impact one and another (Brandl et al., 2024). Recognizing patterns and trajectories in learners' clinical reasoning, including moments where their decision-making deviates from the ideal path, can help tailor instructional support (Banning, 2008). Gaining insight into the clinical decision-making process allows for real-time tracking of learners' reasoning patterns, making it possible to offer timely support and interventions when needed.

Sequential analysis, a method to analyze sequential data and identify meaningful temporal patterns in big data, has been widely used in educational research to examine students' learning behaviors (Baker et al., 2016; Romero & Ventura, 2020). For instance, Li et al. (2020) examined the temporal dynamics of students' self-regulated learning (SRL) behaviors in a computer-aided engineering design learning environment. Temporal analysis allows them to uncover meaningful transitions among different types of SRL learning strategies that potentially contribute to positive learning outcomes. Another study also applied sequential analysis to uncover patterns in transformative and non-transformative discourse in online collaborative learning settings (Zhu et al., 2019). As one of the most used learning analytic methods, sequential analysis has brought many valuable insights into understanding learning behaviors and predicting learning outcomes (e.g., Baker et al., 2016; Du et al., 2024; Saqr & López-Pernas, 2023; Zhu et al., 2019). Yet, medical education research has not fully benefited from this powerful method.

To address this gap, the current study aimed to explore how utilizing sequential analysis can overcome the shortcomings of traditional assessment methods in capturing the dynamic aspects of clinical decision-making. Specifically, we applied Markov Chains transition matrix, a powerful statistical model to examine temporal data, to understand the clinical decision-making process in a medical high-acuity simulation scenario.

 © ISLS

Study design

Participants and setting

The study involved 21 board-certified anesthesiologists (14 males and 7 females) who participated in a simulation-based maintenance of certification course at a Midwestern academic medical center over a five-year period. The simulated learning scenario centered on the intraoperative management of a 36-year-old female patient with acute cholangitis undergoing a laparoscopic cholecystectomy. During the procedure, the patient experienced malignant hyperthermia (MH), requiring the primary anesthesiologist to promptly recognize the condition and initiate appropriate treatment. Established treatment protocols for MH include discontinuing triggering agents, administering dantrolene, and providing supportive care, which are well-documented and widely accessible.

The simulated training scenario involved a team of medical crew, including nurses, surgeon assistants, and surgeons, like real-life surgeon scenarios. The anesthesiologist participants' task as team leaders was to recognize the patient's condition and make a treatment plan for the patient, which is also known as clinical decision-making. They were expected to ask questions about the patient, evaluate the information, state plans, and eventually implement their treatment plan.

Measurement

Team reflection behavioral observation (TuRBO) system

The research team first transcribed the video and then annotated the data using the Team Reflection Behavioral Observation (TuRBO) System which was specifically designed for acute care teams (Schmutz et al., 2021). It serves as a framework to observe and evaluate team reflection during clinical action. TuRBO allows team members to pause amidst dynamic activities to collaboratively collect information, evaluate the current situation, and conduct short-term planning. Researchers examined its validity and suggested that TuRBO effectively assesses teamwork aligning with performance outcomes (Schmutz et al., 2021), as well as provides explicit behavioral indicators for training purposes, particularly in acute care team settings.

We applied TuRBO codes to the transcripts of our video recordings. Two researchers independently coded six data files, then discussed their findings and resolved any discrepancies. They reached an inter-rater reliability of .73 using Cohen's kappa. Following this process, they completed annotating all data files. Then, we combined similar TuRBO codes and summarized a clinical behavioral category for our study. An additional "implementation" category was introduced to account for the critical importance of recognizing malignant hyperthermia (MH) and subsequently implementing appropriate treatment. Table 1 presents the codebook with a few examples from the study.

Table 1
TuRBO with Examples

	TuRBO code	Description	Examples
Seeking information	Actively inviting input	All statements that request information from the team about the current event and invite team members to provide information and share ideas.	"What type of abx did this patient get? Do we need a repeat dose? Can you add abx?"
	Expressing uncertainty	Expressions of uncertainty with an implicit invitation to share information	"Is there anything we are missing? Is there anything else we should be doing?"
Evaluating information	Starting a working hypothesis	Clear formulation of a working hypothesis or diagnosis about the current situation.	"Patient is becoming more tachycardic, do you think there is a lot of bleeding?"
	Recapping	Various pieces of information are brought together, and a summary is provided.	"The patient is in mid-thirties. She's scheduled for lap chole. She started to develop tachycardia and hyperthermia at an undisclosed rate shortly we have given an antibiotic and contacted MH hotline."
	Explicitly assessing the situation	Providing an explicit judgment for something, giving value to a certain process, information or strategy; this can be the process of evaluation information that has been gained through seeking information.	"Paralysis: Patient seems tight, make sure she is paralyzed."

 © ISLS

	Reasoning	Explaining why something is more important than something else or why a specific behavior needs to be done.	"Elevated CO2: this could be due to difficulty ventilating or drug reaction."
Planning	Stating plans and priorities	Laying out the course of action for the next few minutes. Needs to contain at least 2 actions to show a sequence of actions.	"Going to stop the agent and go up on flows."
Implementation		States they are doing the task or delegates a task to another team member.	"Administer dantrolene now."

Performance

Participants were assigned to either a low-performing or a high-performing group based on their performance, which was determined based on three key metrics in medical simulation training: how long it takes to request assistance, how long it takes to ask the surgeon to halt the operation, and how long it takes to administer the initial dose of dantrolene. These three measurements were critical to effectively manage MH during surgeries. If the participant could recognize the patient's situation and administer treatment (dantrolene) within a short period of time, we can expect a better patient outcome. As a result, 16 participants were assigned to the low-performing group, while 5 to the high-performing group.

Data analysis

A Markov Chain transition matrix is a mathematical tool used to represent the probabilities of moving between different states in a Markov process. This process models transitions between states based on predefined probabilistic rules. This approach enables researchers to study the learning process from an event-driven perspective while identifying patterns and transitions within the data (Viberg et al., 2018; Zhu et al., 2019). Markov chain methods have been utilized in educational research to offer sequential recommendations for learning behavior/outcome prediction, course selections, and identifying at risk students (e.g., Beal & Cohen, 2008; Yildiz, 2025). Mathematically, for each state $X_1, X_2, \ldots, X_n$, the probability (Pr) of moving to next state is:

$$Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = Pr\ (X_{n+1} = x | X_n = x_n)$$

For our study, each participant's statement can be seen as an event or a state, represented by the TuRBO code from Table 1. Therefore, we can create a matrix where each element represents the probability of moving from one behavior to another.

Results

To examine the sequential transition between participants' behaviors, we applied a Markov Chain transition matrix for the high-performing and low-performing group, respectively. Results are shown in Table 2 and 3. In these matrices, each row sums to one, represents the total probability of transitioning from a specific behavior to the subsequent behavior. Essentially, the rows correspond to the initial behaviors, while the columns indicate the subsequent behaviors.

Table 2

A Markov Chain Transition Matrix for the Low-Performing Group.

	Seeking info.	Evaluating info.	Planning	Implementation
Seeking info.	.818	.103	.027	.053
Evaluating info.	.698	.202	.057	.042
Planning	.599	.136	.025	.241
Implementation	.672	.085	.040	.202

Table 3

A Markov Cain Transition Matrix for the High-Performing Group.

	Seeking info.	Evaluating info.	Planning	Implementation
Seeking info.	.773	.116	.041	.070
Evaluating info.	.722	.153	.051	.073
Planning	.585	.094	.113	.208
Implementation	.690	.090	.040	.180

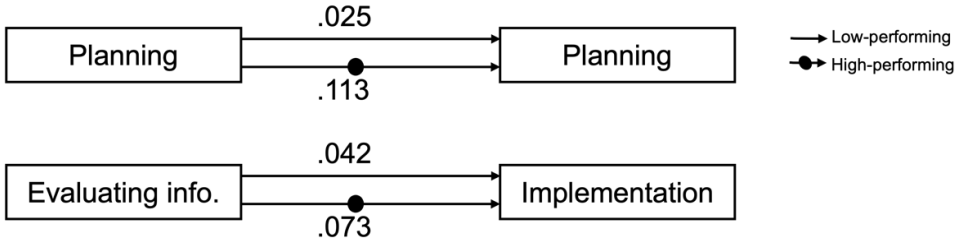
The top three transition pairs for both groups are: seeking information → seeking information, evaluating information → seeking information, and implementation → seeking information. This is logical because, in real-life emergency scenarios, medical professionals must gather as much information as possible to determine the cause and formulate an effective treatment plan. We then conducted a Chi-Square test to analyze the differences

 © ISLS

in transitions between the two groups. While the results indicated a statistically significant difference overall, no specific transition pairs showed statistical significance, especially after applying Bonferroni corrections to adjust for multiple comparisons. We hypothesized that this may be due to a small effect size or the dispersion of differences across numerous transitions between the groups.

We continued to identify transition differences based on transition probability. We defined a notable pair as any transition pair whose transition probability in one group is 50% higher than that in the other group. As a result, we identified two notable pairs: evaluating information → implementation and planning → planning, see Figure 1. For both pairs, the high-performing group has a 70% and 350% higher transition probability than the low-performing group. The big difference in transition probability, especially for the planning → planning pair, highlights the importance of planning to clinical decision-making in our study.

Figure 1
Notable Pairs Based on Transition Probability



Discussion and conclusion

We applied Markov Chain transition matrix to examine the clinical decision-making process data in a simulated medical training setting. Results suggest that the transition matrix of the low-performing group and the high-performing group are statistically significantly different from each other. While the two groups shared similar top three common transition pairs, we identified two notable pairs (evaluating information → implementation and planning → planning) that are different in transition probability. The study has both pedagogical and methodological implications for understanding the clinical decision-making process.

First, both high and low-performing groups frequently transitioned from various states back to seeking information, reflecting the iterative nature of evidence generation during collaborative diagnostic reasoning in high-stakes, time-sensitive scenarios where incomplete or evolving information is common (Salas et al., 2009; Smith et al., 2008). Second, the high-performing group had a 70% higher probability of transiting from evaluating information to implementation, indicating a more efficient information gathering or information integration into their decision-making process. This group seems to better translate diagnostics assessments into therapeutic actions, which is a critical skill in emergency situations where timely intervention significantly impacts patient outcomes. Third, the high-performing group had a 350% higher probability of transiting from planning to planning, highlighting the critical role of planning in clinical decision-making process. We hypothesize that the high-performing group could state their plans and priorities more frequently, contributing to in-time diagnosis and treatment. This also suggests that they engage in more iterative planning, continuously using the input received to formulate a diagnosis and management plan.

Methodologically, Markov-Chain-based method provides a promising approach to understand the dynamic nature of clinical decision-making. We modeled the sequence and transitions between different behavioral states (i.e., seeking information, evaluating information, planning, and implementation) to uncover patterns in decision-making processes that may not be apparent through other traditional assessment methods (Banning, 2008). In our study, Markov Chain transition matrix helped identify both common patterns and notable differences between the high-performing and low-performing groups. In addition, this data-driven approach emphasizing the behaviors exhibited during the simulation could potentially enable in-time and/or post-simulation feedback on these advanced skills.

There are several limitations. First, the limited sample size of this study may impact the statistical power. Researchers and practitioners shall be cautious when generalizing our results to other scenarios. Second, potential confounding factors such as individual differences in prior experience and variations in team composition were not fully accounted for in our analysis, which might influence team performance and decision-making patterns. Future research should extend the current in terms of sample size and learning scenarios to fully leverage the potential of this method in medical education.

 © ISLS

References

Baker, R. S., Martin, T., & Rossi, L. M. (2016). Educational data mining and learning analytics. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, 379-396.

Banning, M. (2008). A review of clinical decision making: models and current research. *Journal of clinical nursing*, *17*(2), 187-195.

Beal, C. R., & Cohen, P. R. (2008). Temporal data mining for educational applications. In *PRICAI 2008: Trends in Artificial Intelligence: 10th Pacific Rim International Conference on Artificial Intelligence, Hanoi, Vietnam, December 15-19, 2008. Proceedings 10* (pp. 66-77). Springer Berlin Heidelberg.

Brandl, L., Stadler, M., Richters, C., Radkowitsch, A., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2024). Collaborative problem-solving in knowledge-rich domains: A multi-study structural equation model. *International Journal of Computer-Supported Collaborative Learning*, *19*(3), 341-368.

Daniel, M., Rencic, J., Durning, S. J., Holmboe, E., Santen, S. A., Lang, V., ... & Gruppen, L. D. (2019). Clinical reasoning assessment methods: a scoping review and practical guidance. *Academic Medicine*, *94*(6), 902-912.

Du, H., Zhu, G., Xing, W., & Xie, C. (2025). What Distinguishes Students' Engineering Design Performance: Design Behaviors, Design Iterations and Application of Science Concepts. *Journal of Science Education and Technology*, *34*, 314-326. https://doi.org/10.1007/s10956-024-10184-y

Gruppen, L. D. (2016). Clinical reasoning: defining it, teaching it, assessing it, studying it. *Western Journal of Emergency Medicine*, *18*(1), 4.

Li, S., Du, H., Xing, W., Zheng, J., Chen, G., & Xie, C. (2020). Examining temporal dynamics of self-regulated learning behaviors in STEM learning: A network approach. *Computers & Education*, *158*, 103987.

Richters, C., Schmidmaier, R., Popov, V., Schredelseker, J., Fischer, F., & Fischer, M. R. (2024). Intervention skills–a neglected field of research in medical education and beyond. *GMS Journal for Medical Education*, *41*(4), Doc48.

Salas, E., Almeida, S. A., Salisbury, M., King, H., Lazzara, E. H., Lyons, R., ... & McQuillan, R. (2009). What are the critical success factors for team training in health care? *The Joint Commission journal on quality and patient safety*, *35*(8), 398-405.

Saqr, M., & López-Pernas, S. (2023). The temporal dynamics of online problem-based learning: Why and when sequence matters. *International Journal of Computer-Supported Collaborative Learning*, *18*(1), 11-37.

Schmidt, H. G., & Boshuizen, H. P. (1993). On the origin of intermediate effects in clinical case recall. *Memory & cognition*, *21*, 338-351.

Schmutz, J. B., Lei, Z., & Eppich, W. J. (2021). Reflection on the fly: development of the team reflection behavioral observation (TuRBO) system for acute care teams. *Academic Medicine*, *96*(9), 1337-1345.

Smith, M., Higgs, J., & Ellis, E. (2008). Factors influencing clinical decision making. *Clinical reasoning in the health professions*, *3*, 89-100.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, *10*(3), e1355.

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in human behavior*, *89*, 98-110.

Yildiz Durak, H. (2025). Impact of ML-LA feedback system on learners' academic performance, engagement and behavioral patterns in online collaborative learning environments: A lag sequential analysis and Markov chain approach. *Education and Information Technologies*, *30*(2), 2623-2644.

Zhu, G., Xing, W., & Popov, V. (2019). Uncovering the sequential patterns in transformative and non-transformative discourse during collaborative inquiry learning. *The Internet and Higher Education*, *41*, 51-61.

Acknowledgments

This work was supported by the National Science Foundation under grant number IIS-2202451. Any opinions, findings, and conclusions or recommendations expressed in this paper, however, are those of the authors and do not necessarily reflect the views of the NSF.