Koh, E., et al. (Eds.) (2026). Proceedings of the 2nd International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education.

Learner Behavior Profiles during Problem Formulation in Problem-based Learning

Deepti REDDY[a*], Tanmay SHARMA[a], Rwitajit MAJUMDAR[b]

[a] *Mukesh Patel School of Technology Management & Engineering, SVKM's NMIMS, Mumbai, India'*

[b] *Research and Education Institute for Semiconductors and Informatics, Kumamoto University, Japan*

Abstract: This study examines student learner behavioral patterns during problem formulation in a Problem-Based Learning (PBL) activity conducted in a micro-learning platform with integrated generative AI support named LA-ReflecT. Students' interaction log data were analyzed to understand how different behavioral strategies relate to the quality of the generated problems. Interaction sequence data from 70 students were modeled using the Longest Common Subsequence (LCSS) similarity measure and clustered using the K-Medoids algorithm. Three distinct behavioral profiles emerged: *Focused and goal-oriented*, *Exploratory and inconsistent*, and *Balanced help-seeking*. Although students exhibiting focused and balanced behaviors achieved slightly higher average problem-quality scores, there were no statistically significant performance differences across the three profile clusters.

Keywords: Problem-based learning, Problem identification, quality of problem, behavior patterns, clustering algorithm.

1. Introduction

Problem-Based Learning (PBL) is a pedagogical approach that enables students to learn by solving real-world problems in a collaborative setting (Yew, 2016). There has been an increasing number of studies examining the effectiveness of PBL on the quality of student learning and the extent to which it delivers on its promise of developing self-directed learning habits, problem-solving skills, and deep disciplinary knowledge (Majoor, 1990). The studies used student reflections and questionnaires to examine the effectiveness of PBL problems in learning. The results showed that the quality of problems had a more direct and stronger influence on the learning outcome, and thereby supported the postulation that a good problem leads to improved learning (Gijselaers and Schmidt, 1990).

Most PBL studies are based on surveys and perceptions; however, only few focus on students' behaviour when selecting and defining their problem statement for PBL in the era of gen-AI. We have conducted a study in which students interacted with the activity on a micro learning platform LA-ReflecT with the aim to help students to define an ill-structured, open-ended problem based on a real-life scenario, having multiple solutions, that requires inquiry and teamwork.

The students' interactions withing LA-ReflecT were logged and used to identify the behavioural patterns exhibited by the students and their impact on the quality of the problem defined. Sequential interaction data from 70 students were modeled using the Longest Common Subsequence (LCSS) similarity measure and clustered using the K-Medoids algorithm. This study seeks to identify behavioural patterns of students when defining their problem statements for PBL, and its impact on the quality of the problem statement.

2. Related work

Learning analytics is used to support and inform PBL practice. In a study by Saleh (2022), PBL was used as a pedagogical framework to study individual and collaborative actions in a game-based learning environment. Principal component analysis (PCA) was used to cluster students and to describe patterns in game interaction data. Sequence-clustering approaches are widely used to group sequences based on similarity measures, using techniques such as optimal matching, k-means, or edit distance. In one such approach, qualitatively coded sequences and optimal matching were used to identify students' strategies in a complex problem-solving context (Eichmann, 2020). In another study (He, 2021), the Longest Common Subsequence (LCS) method is used to identify differences between an expert sequence and a test taker's sequence of actions to solve a particular problem.

Existing studies have established the key characteristics of the problem in PBL, which lead to effective learning outcomes, and have used learning analytics approaches to identify problem-solving behaviors. In this paper, we aim to study students' behaviour when selecting their problem statements in the era of Gen AI using sequence-clustering algorithms.

3. Method

3.1 Intervention

An activity was created in the LA-ReflecT Platform (Majumdar et al. 2023) to help students understand PBL and the problem's characteristics. The reading resources on what PBL is, the steps in PBL, and the characteristics of a problem in PBL, with examples from the domain of distributed computing, were provided. Then students were asked to read, identify their problem statement, self-evaluate, and justify whether their problem satisfies the learnt PBL characteristics. The students were encouraged to interact with an integrated chatbot powered by ChatGPT, to seek feedback on their problem, self-evaluate, and modify if necessary.

3.2 Participants

The study involved 76 third-year Computer Science and Engineering students (58% male, 42% female) enrolled in a four-year engineering program in India. The intervention was part of a 15-week Distributed Computing course. Students interacted individually with the LA-ReflecT activity for approximately 1 hour in a laboratory setting. All clicks and responses were logged within the LA-ReflecT system.

3.3 Interaction log data

The log data consisted of user-id, timestamp, and element-type. The element-type variables and their description are shown in Table 1.

Table 1. *Element-type with description*

Element-type	Description
Instruction-notes	Reading material about PBL
problem-statement	Enter problem statement for PBL
ChatGPT	A chatbot to interact with ChatGPT
Justify	Enter justification on why their problem statement in ill-structured
Self-evaluate	Self-evaluate the problem statement for PBL.

3.4 Sequence Analysis and Clustering

3.4.1. Data preprocessing

For each user, interaction events were grouped chronologically to construct individual interaction sequences. From an initial sample of 76 students, 6 were excluded for insufficient interaction, leaving a final sample of 70. Consecutive duplicate interaction types were also eliminated to avoid self-transitions. Since our analysis and modeling required numerical inputs, Element-type was encoded as a unique integer label: 0-ChatGPT, 1-instruction-notes, 2-justify, 3-problem-stm, 4-self-evaluate. Each user’s interaction sequence was then transformed into a numerical sequence as: User1: [1, 3, 0, 3, 0, 3, 0, 1, 2, 1, 2, 1, 3, 0, 4, ...], User2: [1, 0, 3, 0, 4, 2, 1, 0, 1, 2, 4, 0, 3, 1, 4, 2].

3.4.2 Clustering algorithm

User interaction behavior was represented as encoded categorical sequences of varying lengths. Since these sequences are symbolic, traditional distance measures such as Euclidean distance are unsuitable. Therefore, we used the Longest Common Subsequence

(LCSS) metric, which measures the longest common ordered subsequence between two sequences. The normalized LCSS similarity is defined as:

$$LCSS_{norm}(S_i, S_j) = \frac{LCSS(S_i, S_j)}{min(m, n)}$$

To convert this into a distance metric suitable for clustering, we define:

$$D_{LCSS}(S_i, S_j) = 1 - LCSS_{norm}(S_i, S_j)$$

While LCSS captures structural similarity, it does not account for differences in sequence length, which reflect interaction intensity. To address this, we define a composite distance:

$$D_{combined}(S_i, S_j) = \alpha D_{LCSS}(S_i, S_j) + (1 - \alpha) D_{len}(S_i, S_j)$$

where α=0.7, giving higher weight to structural similarity while incorporating behavioral intensity. A pairwise distance matrix was constructed and used as input for clustering. Due to the non-Euclidean nature of the data, K-Medoids clustering was employed, as it operates directly on distance matrices and is robust to outliers. The optimal number of clusters was determined using the Elbow Method, resulting in K=3.

4. Identified Behaviour Profile with Interaction Sequence Clustering Analysis

To identify behavioral patterns, we analyzed sequence length and dominant interaction elements within each cluster. Three distinct clusters emerged: *Focused and goal-oriented*, *Exploratory and inconsistent*, and *Balanced help-seeking*. The *Focused* cluster showed the shortest average sequence length of 9.4, while the *Exploratory* cluster had the longest of 18 and *Balance help-seeking* had length 13.5.

Dominant interactions at each step were identified using normalized frequency distributions, selecting the most probable action per step. The spider chart (Figure 1) highlights distinct patterns: the focused cluster shows consistent early engagement across key activities: reading, editing, use of ChatGPT, followed by the decline in later steps; the exploratory cluster exhibits strong dominance in initially reading and editing problem statements but lacks consistency later; and the balanced cluster demonstrates a strong initial focus on reading and editing the problem statement, with steady use of ChatGPT help and moderate self-evaluation toward the end.

To examine performance, student submissions were scored on a 1–3 scale. A score of 1 indicates that none of the PBL criteria were satisfied, whereas a score of 3 indicates that all criteria were met. The focused cluster achieved the highest average score, while the exploratory cluster had the lowest (Table 2). However, ANOVA results indicated no statistically significant differences between clusters ($p = 0.95$).

Table 2. *Descriptive statistics of student scores by cluster*

	Focused and goal-oriented	Exploratory and inconsistent	Balanced help-seeking
Average score	2.28	2.23	2.27
Std Dev	0.67	0.69	0.64

5. Discussion

This study examined students' behavioral patterns while defining problem statements in PBL and their relation to problem quality. Three distinct clusters were identified. The focused and goal-oriented cluster showed the shortest sequences, reflecting efficient and task-directed behavior. The exploratory and inconsistent cluster had the longest sequences, with initial engagement but lack of consistency in later stages, indicating absence of a clear strategy. The balanced help-seeking cluster demonstrated moderate sequence length, combining structured engagement with sustained use of support tools. Overall, focused and balanced behaviors were associated with slightly better performance than exploratory behavior; however, differences were not statistically significant.

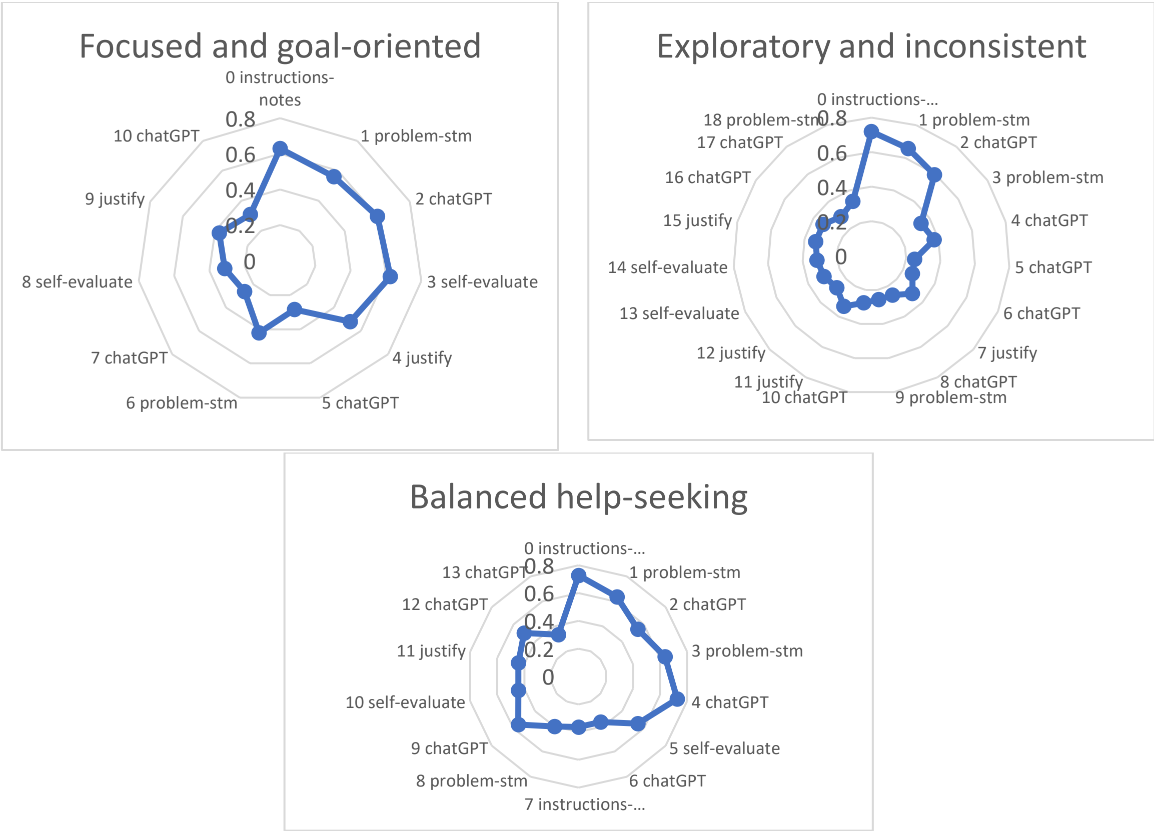

Figure 1. *Behavioral patterns of clusters.*

6. Conclusion

This study analyzed students' interaction patterns in a PBL environment supported by LMS and generative AI. Using LCSS and K-Medoids clustering, three behavioral profiles: *Focused*, *Exploratory*, and *Balanced*, were identified. The findings highlight variations in engagement strategies during problem formulation and their relationship with learning outcomes.

References

Eichmann, B., Greiff, S., Naumann, J., Brandhuber, L., & Goldhammer, F. (2020). Exploring behavioural patterns during complex problem-solving. *Journal of Computer Assisted Learning*, 36(6), 933‑956.

Gijselaers, W., & Schmidt, H. (1990). Development and evaluation of a causal model of problem-based learning.

He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education*, *166*, 104170.

Majoor, G. D., Schmidt, H. G., Snellen-Balendong, H., Moust, J. C. H. & Stalenhoef-Halling, B. (1990) Construction of problems for problem-based learning. In Z. Nooman, H. G. Schmidt & E.S. Ezzat (Eds.), Innovation (pp. 114-122). New York: Springer

Majumdar, R., Prasad, P., Kadam, K., Gatare, K., Warriem, J.M. (2023). LA-ReflecT: A Platform Facilitating Micro-learning and Its Multimodal Learning Analytics. In *Lecture Notes in Computer Science*, vol 14200. Springer, Cham.

Saleh, A., Phillips, T. M., Hmelo-Silver, C. E., Glazewski, K. D., Mott, B. W., & Lester, J. C. (2022). A learning analytics approach towards understanding collaborative inquiry in a problem-based learning environment. British Journal of Educational Technology, 53(5), 1321–1342.

Sockalingam, N., Rotgans, J., & Schmidt, H. (2012). Assessing the Quality of Problems in Problem-Based Learning. *Int. Journal of Teaching and Learning in Higher Education*, *24*(1), 43-51.

Yew, E. H., & Goh, K. (2016). Problem-based learning: An overview of its process and impact on learning. *Health professions education*, 2(2), 75-79.