

A Review of Soft Actor-Critic Methods for Adaptive Multi-Objective Reward Learning

Ashu Tiwari, Savan Payasi, Akhilesh A. Wao*

Department of Computer Science, AKS University, SATNA, MP, India

ABSTRACT

The rapid growth of online learning has increased the need for systems that provide personalized support according to students' knowledge, learning speed, and engagement. Adaptive learning systems address the limitations of fixed learning sequences by using student data to provide suitable learning activities. Knowledge Tracing is widely used to estimate students' changing knowledge levels from previous learning interactions. However, knowledge prediction alone cannot determine which activity should be given next. Reinforcement Learning can address this problem by making sequential decisions based on the learner's changing state. Recent studies have combined Knowledge Tracing with Reinforcement Learning for personalized learning-path recommendation and have also explored Soft Actor-Critic (SAC) for educational recommendation. However, reward design remains a challenge because rewards based only on correctness or short-term performance may favor easy activities rather than meaningful learning improvement. Multi-objective Reinforcement Learning can consider multiple learning goals during decision-making. This review examines Knowledge Tracing, learner-state estimation, Reinforcement Learning, SAC, personalized learning-path recommendation, and multi-objective reward design. Based on the literature, a knowledge-level-aware adaptive learning framework is proposed in which the changing knowledge state guides SAC-based task selection, while the reward considers performance, knowledge gain, and engagement. The review also identifies gaps in reward design, learner-state representation, action-space design, evaluation quality, and long-term educational validation.

Keywords: Adaptive Learning, Knowledge Tracing, Multi-Objective Reward Optimization, Personalized Learning, Reinforcement Learning, Soft Actor-Critic, Student Knowledge State.

1. INTRODUCTION

The growing use of digital education has increased the need for learning systems that can respond to individual student needs. Students differ in their knowledge, progress, learning speed, and engagement, so the same learning path may not be suitable for everyone. Adaptive learning addresses these differences by using student information to provide suitable learning activities. Knowledge Tracing can estimate changes in student knowledge, while Reinforcement Learning, particularly Soft Actor-Critic (SAC), can support the selection of the next learning task. In addition, a multi-objective reward can evaluate learning using performance,

knowledge gain, and engagement. Therefore, this research focuses on these three connected areas.

1.1 Digital Learning and the Need for Adaptive Learning

Digital learning platforms collect student information such as answers, marks, attempts, time spent, learning history, and engagement. This information can help researchers understand student learning behavior and provide support according to individual learning needs (Hariyanto et al., 2025). However, many traditional systems still provide the same sequence of content to students even when their needs are different. Adaptive learning addresses this problem by selecting suitable learning activities according to student progress and needs.

Student knowledge should not be treated as fixed because it changes through study, practice, mistakes, and feedback. Therefore, learning systems should regularly update their understanding of student knowledge instead of depending only on a single score or fixed knowledge category (Qian et al., 2024). Knowledge Tracing supports this process by using previous student interactions to estimate changes in knowledge over time. Recent studies have also considered longer learning histories and additional learner information for better knowledge estimation (Luo et al., 2024; Qian et al., 2024).

1.2 Knowledge Tracing and Reinforcement Learning for Learning Decisions

Knowledge Tracing can estimate what a student currently knows, but it does not decide what the student should study next. Reinforcement Learning can support this decision by selecting learning activities, observing student responses, and learning from the results of previous decisions (Riedmann et al., 2025).

Recent research has also explored SAC for educational recommendation. Song et al. (2026) used a SAC-based approach for personalized educational-resource recommendation based on learner behavior and available resources. SAC can help an adaptive system explore different learning activities rather than repeatedly selecting only familiar tasks. However, the usefulness of a task should not be evaluated only through correct answers. Otherwise, the system may prefer easy questions that improve short-term performance without producing sufficient knowledge improvement (Fu, 2025).

1.3 Multi-Objective Reward and Proposed Research Direction

Learning involves several goals, including improving performance and knowledge while maintaining student engagement. A multi-objective

reward can consider these outcomes together. In this research, performance, knowledge gain, and engagement are the main reward components. Recent studies also support considering multiple learning goals. For example, Yi et al. (2026) considered learning gain, knowledge coverage, and changes in learning difficulty while generating adaptive learning paths.

Although progress has been made in Knowledge Tracing, Reinforcement Learning, SAC-based recommendation, and adaptive learning paths, there is still a need to connect these areas within one adaptive learning framework (Fu, 2025; Li et al., 2026). Therefore, this research focuses on Adaptive Learning Using Soft Actor-Critic with Multi-Objective Reward Optimization. Knowledge Tracing will estimate the changing student knowledge state, SAC will select the next learning task, and a multi-objective reward based on performance, knowledge gain, and engagement will evaluate the response. The updated knowledge state and learning results will then support the next task selection, allowing the learning path to change according to student progress and needs.

2. Review of Literature

Recent research in adaptive learning focuses on understanding the changing knowledge of students and using this information to provide suitable learning activities. The literature related to the present research can be divided into five main areas: Knowledge Tracing and student knowledge modelling, Knowledge Tracing with Reinforcement Learning, Reinforcement Learning and Actor-Critic methods, multi-objective adaptive learning, and personalized adaptive learning.

2.1 Knowledge Tracing and Student Knowledge Modelling

Knowledge Tracing is used to understand how student knowledge changes during learning. Qian et

al. (2024) proposed a student-state-based Knowledge Tracing method that used previous learning information to improve the prediction of student performance.

Luo et al. (2024) also studied student states and showed that different levels of learning information can improve knowledge estimation.

Tong and Ren (2025) combined Knowledge Tracing with cognitive-load information for personalized learning-path generation. Their work showed that correct and incorrect answers alone may not provide enough information about a student's learning condition.

Li et al. (2026) developed an Adaptive Knowledge Tracing approach that considers changes in knowledge, previous learning information, and forgetting. The model was tested on three Assessment's datasets and showed improvements over comparison methods.

Xiang and Li (2026) further considered changing learner information for personalized learning-path optimization. Together, these studies show that student knowledge should be updated during learning rather than treated as a fixed value.

Recent studies show that understanding changes in student knowledge is important for personalized learning. Li et al. (2026) showed that student knowledge can be updated using learning history and forgetting information, while Xiang and Li (2026) found that changing learner information can support personalized learning paths. Tong and Ren (2025) showed that combining student knowledge with other learner information can improve learning-path decisions. Similarly, Qian et al. (2024) and Luo et al. (2024) found that previous student interactions and student-state information can improve the estimation of student knowledge. However, most of these studies mainly focus on understanding or predicting student knowledge and

do not clearly explain how this updated knowledge should be used to select the next learning activity. Therefore, there is a need for an adaptive learning approach that connects the changing student knowledge state with the selection of the next suitable learning task.

2.2 Knowledge Tracing with Reinforcement Learning

Recent studies have started connecting Knowledge Tracing with Reinforcement Learning so that student knowledge can be used directly when selecting future learning activities. Fu (2025) combined Dynamic Knowledge Tracing with Reinforcement Learning for personalized learning-path optimization. The model updated student knowledge and used this information to select learning tasks. Ozyurt et al. (2025) developed an exercise recommendation approach that combined Knowledge Tracing, learning-content information, and different Reinforcement Learning methods. Their experiments showed that knowledge information can support personalized exercise recommendation. Li et al. (2026) also connected changing knowledge information with Reinforcement Learning to improve knowledge-state adjustment. Zhao et al. (2026) proposed a personalized learning-path system that combines Knowledge Tracing with Deep Reinforcement Learning and considers several learning goals. Dong and Zhao (2026) also integrated Knowledge Tracing and Reinforcement Learning within a personalized educational system. These studies show a growing connection between student knowledge estimation and learning decisions.

Recent studies explain how Knowledge Tracing and Reinforcement Learning can be combined to provide learning activities based on student needs. Zhao et al. (2026) used changing student knowledge with Reinforcement Learning to create personalized learning paths, but their model has many parts,

making it difficult to study each reward factor separately. Li et al. (2026) used Reinforcement Learning to update student knowledge during learning, but their main aim was to improve Knowledge Tracing rather than select the next task using SAC. Dong and Zhao (2026) connected student knowledge with learning decisions, but their work was limited to a particular educational setting. Fu (2025) used changing student knowledge for task selection and reported better learning-path results, but the reward method still needs improvement. Ozyurt et al. (2025) used student knowledge and exercise information to select suitable exercises, but their study mainly covered online mathematics activities. Overall, these studies suggest that Knowledge Tracing and Reinforcement Learning can work together for personalized learning. However, more research is needed to develop a model where the changing student knowledge helps select the next suitable task and the reward considers performance, knowledge gain, and engagement.

2.3 Reinforcement Learning and Soft Actor-Critic in Adaptive Learning

Reinforcement Learning is increasingly being used to select learning activities according to student progress. Riedmann et al. (2025) reviewed 89 studies and found that Reinforcement Learning has useful applications in personalized education, including content selection, learning sequences, and feedback. However, the review also identified the need for stronger evaluation and more testing with real learners. Ruan and Lu (2025) used Deep Reinforcement Learning to change personalized learning paths according to student performance and other learning information. Fu (2025) showed that Reinforcement Learning can select tasks using an updated knowledge state. Chunqin and Peixi (2026) used an Actor-Critic approach for course recommendation, showing that this family of methods can be applied in education. Zhao et al.

(2026) further demonstrated the use of Reinforcement Learning for changing learning-path decisions.

Current research has explored the use of Reinforcement Learning in adaptive education for different learning decisions. Chunqin and Peixi (2026) applied an Actor-Critic method for personalized course recommendation, but their work did not focus on selecting the next learning task using Knowledge Tracing. Zhao et al. (2026) used Reinforcement Learning to change learning paths based on student progress, although their method includes several parts and does not mainly examine SAC. Fu (2025) applied Reinforcement Learning to select tasks according to changes in student knowledge, but the study used a simpler learning and reward method than the proposed SAC approach. Riedmann et al. (2025) reviewed Reinforcement Learning in education and reported its value for personalized learning and adaptive decisions. However, many existing approaches still need evaluation with larger datasets and real students. Ruan and Lu (2025) used current learner information to adjust learning-path recommendations, but their method requires different types of student information that may not be available in every dataset. Overall, Reinforcement Learning is becoming useful in adaptive education, but research directly using Soft Actor-Critic for selecting the next learning task is still limited. Therefore, SAC needs to be studied using the changing student knowledge state and a reward that represents important learning outcomes.

2.4 Multi-Factor and Multi-Objective Adaptive Learning

A learning system should not judge the usefulness of a task only by whether the student answers correctly. Ruan and Lu (2025) considered performance, emotion, cognition, and interaction information while changing personalized learning

paths. Fu (2025) considered learning performance and engagement but also reported that its reward design could be improved. Zhao et al. (2026) considered several educational goals while generating learning paths. Yi et al. (2026) developed an adaptive learning-path approach that considered immediate learning improvement, wider knowledge coverage, and changes in learning difficulty. Ozyurt et al. (2025) also considered cumulative knowledge improvement when recommending exercises. Together, these studies indicate that several aspects of student learning should be considered when deciding whether a learning activity is useful.

Previous research has examined various learning factors to enhance adaptive learning for students. Zhao et al. (2026) incorporated knowledge development, exploration, and learning difficulties into learning-path planning; however, their approach utilized multiple objectives rather than the three specific factors selected in this study. Yi et al. (2026) considered learning gain, knowledge coverage, and difficulty levels when constructing learning paths, but their work relies on a reinforcement learning method distinct from SAC. Ruan and Lu (2025) integrated performance, emotion, and cognition into learning-related decision-making; however, this approach requires extensive student data that may not be available in every dataset. Fu (2025) took performance and engagement into account for task selection, yet the reward mechanism requires further refinement. Ozurt et al. (2025) included knowledge improvement and exercise-related information when recommending appropriate exercises, but they did not simultaneously evaluate performance, knowledge gain, and engagement. Overall, prior research supports the importance of utilizing multiple learning outcomes. Nevertheless, there remains a need to investigate a clear reward mechanism based on student performance,

knowledge gain, and engagement, and to understand how these three factors collectively influence SAC-based next-task selection.

2.5 Artificial Intelligence and Personalized Adaptive Learning

Personalized adaptive learning aims to provide learning support according to the needs and progress of individual students. Hariyanto et al. (2025) reviewed 142 studies and found that recent adaptive learning research uses student information to improve personalization, engagement, and learning support. Zhang et al. (2024) developed personalized learning-path recommendation using learning-process information and Deep Knowledge Tracing. Ruan and Lu (2025) showed that learning paths can be changed in real time using learner feedback. Tong and Ren (2025) combined knowledge and cognitive information for personalized learning-path generation. Yi et al. (2026) used changing learner information to generate adaptive learning paths. These studies show that personalized learning is moving from fixed learning sequences toward learning paths that can change as the student progresses.

Research in adaptive learning has explored various methods to provide appropriate learning paths for students. Yi et al. (2026) utilized dynamic learner information to construct personalized learning paths but did not employ SAC as the primary method for task selection. Tong and Ren (2025) integrated student data with other learner information to optimize learning paths, though their approach required additional data such as cognitive load. Ruan and Lu (2025) leveraged student feedback to update learning recommendations during the learning process, yet their model relied on a wide range of learner data. Hariyanto et al. (2025) discussed the role of AI-based methods in personalization, student engagement, and adaptive education, while also highlighting challenges

related to evaluation and practical implementation. Zhang et al. (2024) utilized learning-process information and knowledge tracing for personalized learning paths, but their work did not investigate SAC or the proposed three-component reward mechanism. Overall, these studies establish a solid foundation for personalized and adaptive learning. However, there remains a need for a clear framework that integrates evolving student knowledge, adaptive task selection, educational rewards, and updated learning decisions into a quantifiable learning process.

2.6 Overall Findings and Difference of the Study

Table 1: Summary of Research Gaps Identified from the Literature and the Direction of the Present Study

Area	What Current Research Provides	Remaining Gap	Direction of Present Research
Knowledge Tracing	Estimates changing student knowledge from previous interactions.	Mainly predicts knowledge; does not itself select the next task.	Use the changing knowledge state for task selection.
KT + Reinforcement Learning	Connects student knowledge with learning decisions.	Reward design and clear educational objectives require further study.	Connect Knowledge Tracing directly with SAC-based task selection.
Reinforcement Learning	Supports repeated and personalized learning decisions.	Direct educational use of SAC for next-task selection remains limited.	Study Soft Actor-Critic as the main adaptive learning agent.
Multi-Objective Learning	Shows the value of considering several learning outcomes.	The combined effect of performance, knowledge gain, and engagement needs focused evaluation.	Build and test a reward using these three factors.
Personalized Adaptive Learning	Provides learning activities according to learner needs and progress.	Complete frameworks connecting knowledge change, task selection, and reward evaluation remain an active research need.	Develop an adaptive learning process that continuously updates student knowledge and recommendations.

The main research gap identified from these categories is the need for a clear adaptive learning approach that connects a student's changing knowledge state with Soft Actor-Critic-based next-task selection and a multi-objective reward based on performance, knowledge gain, and engagement. Therefore, the proposed research, “Adaptive Learning Using Soft Actor-Critic with Multi-

The reviewed literature shows that recent research has made important progress in understanding changing student knowledge, using Knowledge Tracing for learner modelling, applying Reinforcement Learning to educational decisions, and developing personalized learning paths. Knowledge Tracing can provide an updated view of what a student knows, while Reinforcement Learning can use this information to select future learning activities. Recent studies have also shown that considering several learning outcomes can provide a better basis for adaptive decisions than using only correct and incorrect answers.

Objective Reward Optimization,” will use Knowledge Tracing to estimate the changing student knowledge state. This information will be provided to the SAC agent to select the next learning task. After the student responds, performance, knowledge gain, and engagement will be used to calculate the reward. The updated student knowledge and learning results will then support the next task

decision. The proposed model will also be compared with suitable baseline methods and evaluated for students with different starting knowledge levels.

3. RESEARCH GAP

Based on the review of recent studies, the following major research gaps have been identified:

1. Existing methods do not effectively use students' changing knowledge levels to select suitable learning tasks.

Knowledge tracing methods can estimate how student knowledge changes over time, and recent studies have started combining Knowledge Tracing with Reinforcement Learning. However, there is still a need for a clear framework where the changing knowledge state of a student is directly used to select the most suitable next learning activity. The updated knowledge state can provide useful information about the student's current learning needs. This information can help the system decide whether the student needs practice, revision, or a more challenging task. The next activity should change as the student's knowledge improves. This connection can make task selection more suitable for individual students.

2. Limited Multi-Objective Reward Design

Many reinforcement learning-based educational systems use relatively simple reward functions. A reward based mainly on correctness or immediate performance may not fully represent actual student learning. There is a need to design and evaluate a reward function that considers student performance, knowledge gain, and engagement together. These three factors represent different parts of the learning process. Performance shows how well the student completes a task, while knowledge gain shows improvement after learning. Engagement indicates how actively the student takes part in the learning activity. Considering these factors together may

provide a better way to evaluate the value of a selected task.

3. Need for Knowledge-Level-Aware Personalized Learning

Students with Low, Medium, and High knowledge levels may require different types and difficulty levels of learning activities. However, the effectiveness of an adaptive policy should be studied separately for different learner groups rather than reporting only one overall result. Recent research also emphasizes dynamically adapting learning paths according to evolving student knowledge. Students with different starting knowledge levels may respond differently to the same learning activity. Therefore, separate evaluation can help understand how the model works for each student group. It can also show whether the selected tasks match the student's current learning needs. This can provide a clearer understanding of personalized learning for students with different knowledge levels.

4. Need for Strong Evaluation and Comparison of Reinforcement Learning Models

Existing educational RL research still has limitations related to evaluation methods, baseline comparisons, statistical testing, and large-scale testing with real learners. Therefore, a proposed SAC-based adaptive model should be compared with suitable methods such as DQN, PPO, A2C, rule-based recommendation, and Knowledge Tracing without RL using multiple educational evaluation measures. Such comparisons can show whether SAC provides better learning decisions than other methods. Using different evaluation measures can provide a more complete view of model performance. The comparison can also identify the strengths and limitations of each method. This will help determine whether the

proposed model provides meaningful improvement over existing approaches.

5. Limited Study of an Interpretable Knowledge-Level-Aware SAC Framework

Recent studies already combine Knowledge Tracing, Reinforcement Learning, and multiple learning objectives, so simply combining these techniques cannot be considered a sufficient research gap. There is still scope to study a simpler and more interpretable framework in which the learner's changing knowledge state guides Soft Actor-Critic-based task selection, while the separate effects of performance, knowledge gain, and engagement are clearly evaluated. A clear framework can help explain how student knowledge is used when selecting the next learning task. It can also show how each reward factor affects the decisions made by the SAC model. Studying these factors separately can provide a better understanding of their role in adaptive learning. This can make the results easier to understand, compare, and evaluate.

4. Research Methodology

The methodology of this study was designed in two phases. The first phase focused on the analysis of multiple educational datasets to understand different types of student information and identify a suitable dataset for further adaptive learning research. Five educational datasets were considered: the College Student Behavioral Dataset, EdNet-KT1, Learning Preferences Dataset, AI Course Performance Dataset, and SPHERE Dataset. These datasets provide different forms of information, including academic performance, student behavior, learning preferences, assessment results, and sequential learning interactions. Each dataset was examined separately because the available features and the meaning of student performance were different across the datasets.

4.1 Data Preprocessing and Feature Selection

The datasets were first examined for their structure, available features, missing values, duplicate records, and data types. Data cleaning was performed according to the requirements of each dataset rather than applying the same cleaning method to all datasets. Relevant features were then selected for student analysis. Academic measures such as GPA, course grades, quiz scores, final examination scores, and test performance were used where available. For EdNet-KT1, sequential student-question interactions and official answer information were used to examine student responses across learning activities. The Learning Preferences Dataset was mainly considered for learner-profile information because its available features do not provide a direct measure of student knowledge.

4.2 Dataset-Specific Student Analysis

A dataset-specific approach was used to represent student performance because the five datasets contain different educational information. In the College Student Behavioral Dataset, academic information such as GPA and course grades was used. The AI Course Performance Dataset used quiz and final-examination information, while the SPHERE Dataset used final-test performance and other learner attributes. EdNet-KT1 was analyzed using sequential student-question interactions and answer correctness. The Learning Preferences Dataset was treated mainly as a source of learner-profile information rather than as direct evidence of knowledge. This approach avoided using one common score for datasets that measure different aspects of student learning.

4.3 Student Group Classification

Where a suitable student performance or knowledge score was available, students were divided into Low, Medium, and High groups using rank-based tertile classification. This method was used to create comparable groups for descriptive analysis.

Therefore, the nearly equal number of students in the three groups is a result of the classification method and should not be interpreted as the natural distribution of student knowledge in the population. These groups represent the students' observed performance in the completed analysis and are not considered permanent knowledge states.

4.4 Descriptive and Comparative Analysis

Descriptive analysis was performed to understand the distribution of student scores and other important educational features within each dataset. The Low, Medium, and High groups were compared to examine differences in student performance and related learning information. The datasets were also compared according to their size, available features, type of educational information, and suitability for further adaptive learning research. This comparison showed that each dataset contributes a different type of information and that a single dataset cannot represent every aspect of student learning.

4.4.1 Comparative Summary of the Five Datasets

Table 2 presents the number of students, number of columns, and the Low, Medium, and High group sizes obtained for each dataset after the

classification step described in Section 4.3. The number of students analyzed ranged from 144 in the Learning Preferences Dataset to 5,000 in EdNet-KT1, showing a large difference in scale among the five datasets (Figure 3).

Table 3 presents the mean, standard deviation, minimum, and maximum knowledge or performance score obtained for each dataset, where such a score was available. The College Student Behavioral Dataset recorded the highest mean score (0.7831), followed by EdNet-KT1 (0.6177) and SPHERE (0.5751), while the Learning Preferences Dataset recorded the lowest mean score (0.1891), consistent with its use of a provisional, age-based proxy rather than a direct performance measure (Figure 4). A mean score for the AI Course Performance Dataset was not available in the saved output and is therefore not included in this comparison.

These results support the observation in Section 4.4 that each dataset contributes a different type of educational information, and that no single dataset was sufficient to represent every aspect of student learning required for the proposed research.

Table 2: Dataset Summary

Dataset	Students	Columns	Low	Medium	High
College Student Behavioral	1,439	17	480	479	480
EdNet-KT1	5,000	Not reported	1,667	1,666	1,667
Learning Preferences	144	106	48	48	48
AI Course Performance	410	16	137	136	137
SPHERE	497	335	166	165	166

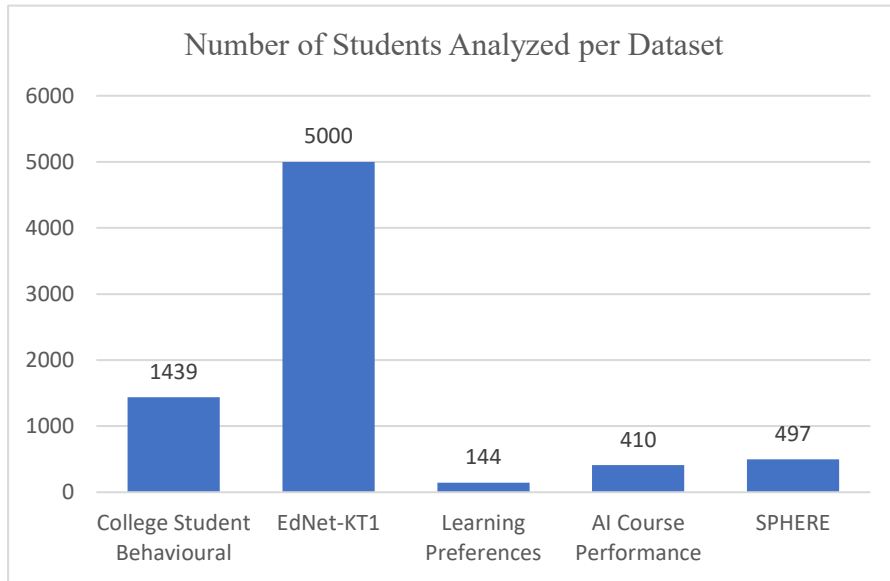


Figure 3: Number of Students Analysed per Dataset

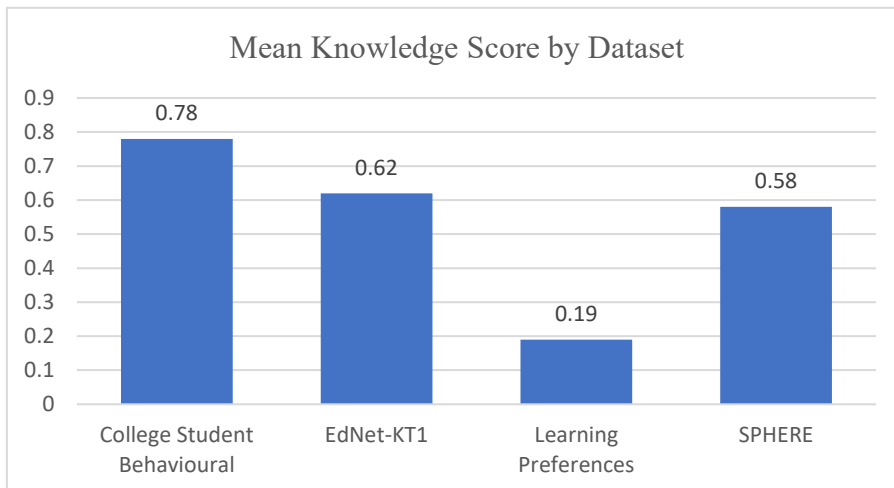


Figure 4: Mean Knowledge Score by Dataset

4.5 Selection of EdNet-KT1 for Sequential Modelling

Among the five datasets, EdNet-KT1 was selected for the next stage of the research because it provides detailed sequential learning interactions. The analyzed EdNet-KT1 data contains 5,000 students and 6,225,147 interaction records, with official answer information available for all analyzed interactions. Unlike datasets that mainly provide final academic or assessment results, EdNet-KT1 records a sequence of student-question interactions. This makes it suitable for studying how student knowledge changes over time and for developing a model that can use previous interactions when selecting future learning activities.

4.6 Proposed Adaptive Learning Framework

Based on the multi-dataset analysis and findings from the reviewed literature, the next phase proposes an adaptive learning framework using Knowledge Tracing and Soft Actor-Critic. Knowledge Tracing will estimate the changing knowledge state of a student from sequential interactions in EdNet-KT1. The estimated knowledge state, together with relevant learning information, will form the learner state provided to the SAC agent. The SAC

agent will use this state to select the next suitable learning task. After the student attempts the selected task, the response will be used to update the knowledge state and calculate a multi-objective reward.

The proposed reward will consider three main factors: student performance, knowledge gain, and engagement. Different combinations and levels of importance for these factors will be examined to understand their effect on task selection. The SAC-based approach will later be compared with suitable baseline methods, including Knowledge Tracing without Reinforcement Learning, DQN, PPO, and A2C. Single-objective and multi-objective reward methods will also be compared. The proposed model will further be evaluated for students with different starting knowledge levels. Since the proposed SAC model has not yet been trained and experimentally evaluated, this stage represents the planned methodology rather than completed experimental results.

4.7 Methodology Flow

The overall methodology follows a step-by-step process from multi-dataset analysis to the proposed adaptive learning framework. The flow presents both the completed student data analysis and the proposed use of Knowledge Tracing, Soft Actor-Critic, and Multi-Objective Reward Optimization for adaptive learning.

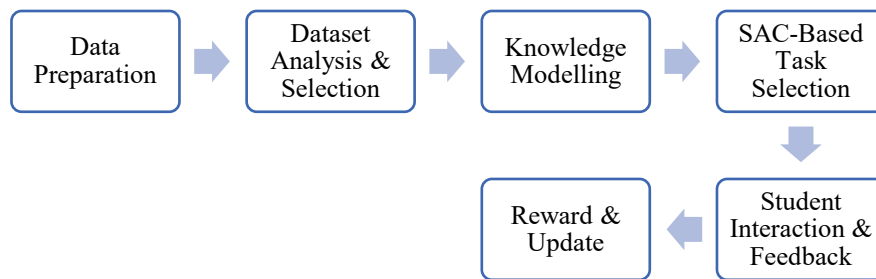


Figure 1: Methodology Flow of the Proposed Adaptive Learning Framework

6. DISCUSSION

The reviewed studies show important progress in student knowledge modelling, personalized learning, and learning-path recommendation. Knowledge tracing estimates changes in student knowledge, while Reinforcement Learning supports the selection of suitable learning activities. However, important gaps remain, as many approaches use limited reward measures or mainly focus on student performance. The combined use of student performance, knowledge gain, and engagement, along with SAC-based task selection using a changing knowledge state, requires further study. The analysis of five educational datasets provides a foundation for addressing these gaps. Among them, EdNet-KT1 is suitable for the next stage because the analyzed data contains 5,000 students and 6,225,147 sequential interactions with

complete answer information. Based on these findings, the next stage will develop an adaptive learning framework using Soft Actor-Critic with Multi-Objective Reward Optimization. Knowledge Tracing will update the student's knowledge state after each interaction, and SAC will use the updated state to select the next suitable learning task. The reward will combine student performance, knowledge gain, and engagement. The proposed approach will be compared with Knowledge Tracing without Reinforcement Learning, DQN, PPO, A2C, and single-objective reward methods. It will also be evaluated for different starting knowledge levels, with the long-term aim of continuously adapting learning activities to each student's progress and needs.

7. CONCLUSION

This study provides a data-based foundation for the proposed Adaptive Learning Using Soft Actor-Critic with Multi-Objective Reward Optimization. In the completed first phase, five educational datasets were analyzed to study student knowledge, performance, behavior, learning preferences, and sequential interactions. These included 1,439 students from the College Student Behavioral Dataset, 5,000 from EdNet-KT1, 144 records from the Learning Preferences Dataset, 410 students from the AI Course Performance Dataset, and 497 students from the SPHERE Dataset. Students were classified into Low, Medium, and High groups using rank-based tertile classification; therefore, the nearly equal group sizes result from the classification method rather than the natural distribution of knowledge levels. The datasets provide different educational information. The College Student Behavioral, AI Course Performance, and SPHERE datasets provide academic performance information, while the Learning Preferences Dataset mainly provides learner-profile information. Among these datasets, EdNet-KT1 provides the strongest basis for sequential modelling, with 5,000 students and 6,225,147 interactions and complete answer metadata.

Based on these findings, the next phase will use Knowledge Tracing to estimate students' changing knowledge states. This learner state will be provided to a Soft Actor-Critic (SAC) agent for adaptive learning task selection. A multi-objective reward combining student performance, knowledge gain, and engagement will guide task selection and support the learning process. The proposed model will be compared with suitable baseline approaches and evaluated across different learner groups. Since SAC training and evaluation have not yet been completed, no experimental performance claim is made at this stage. Overall, the completed analysis

provides a foundation for connecting changing student knowledge with SAC-based adaptive task selection and multi-objective reward optimization, allowing future learning tasks to be selected according to students' changing learning needs and progress.

REFERENCES

- [1]. Li, L., Duan, Z., Zhou, Z., & Liu, L. (2026). Adaptive knowledge tracing with dynamic memory and reinforcement learning. **Sensors*, 26*(6), 1878. <https://doi.org/10.3390/s26061878>
- [2]. Song, Z., Geng, L., Wang, X., & Gong, X. (2026). Deep reinforcement learning-based resource recommendation system for ideological and political education. **Scientific Reports**. <https://doi.org/10.1038/s41598-026-63860-x>
- [3]. Yi, F., Zhu, D., & Li, S. (2026). Adaptive knowledge graph and learning path generation algorithm for artificial intelligence. **Discover Artificial Intelligence*, 6*, 722. <https://doi.org/10.1007/s44163-026-01379-w>
- [4]. Shimada, D., & Okada, K. (2026). Reliability coefficient for Bayesian knowledge tracing models. *Technology, Knowledge and Learning*, 31, 585–604. doi:10.1007/s10758-025-09829-7
- [5]. Li, Y., Zhou, T., Cai, T., & Ju, S. (2026). Interpretable knowledge tracing with dual-level knowledge states. *Expert Systems with Applications*, 298, 129658. doi:10.1016/j.eswa.2025.129658
- [6]. Yin, J., Qiao, J., Xu, H., Goh, T.-T., & Hu, Y. (2026). Metacognitive skills driven knowledge tracing. *Expert Systems with*

- Applications, 297, 129229.
doi:10.1016/j.eswa.2025.129229
- [7]. Liu, W., Yang, B., Su, H., Wang, Y., & Li, Q. (2026). EKT-ML: An efficient knowledge tracing model with multi-task learning. Expert Systems with Applications, 298, 129671.
doi:10.1016/j.eswa.2025.129671
- [8]. Liu, G., Wang, L., Li, J., Li, X., Zhang, X., & Zhang, S. (2026). PG-LLM-KT: An adaptable prerequisite-guided framework for knowledge tracing with frozen lightweight LLMs. Knowledge-Based Systems. doi:10.1016/j.knosys.2026.116926
- [9]. Ni, E., Wang, J., Luan, J., Liu, J., Xiong, S., Xu, S., Cheng, W., Chen, M., Ni, Y., & Li, Y. (2026). Interpretable knowledge tracing via explicit–implicit alignment. Knowledge-Based Systems, 344, 116071.
doi:10.1016/j.knosys.2026.116071
- [10]. Wang, S., Qiu, J., Pang, J., Yang, X., Xu, K., Duan, Z., Zhong, Y., & Zhang, X. (2026). Towards personalized long-term learning modeling in knowledge tracing. Expert Systems with Applications, 331, 133212. doi:10.1016/j.eswa.2026.133212
- [11]. Yu, G., Zhao, Z., Zhou, G., Xie, Z., & Shen, J. (2026). Attributes enhanced representation learning with large language model for cold-start knowledge tracing. Expert Systems with Applications, 319, 131985. doi:10.1016/j.eswa.2026.131985
- [12]. Qiu, L., & Zhu, Q. (2026). A temporal interaction-enhanced multi-view framework for knowledge tracing via self-supervised contrastive learning. Knowledge-Based Systems, 337, 115332.
doi:10.1016/j.knosys.2026.115332
- [13]. Yang, H., Hu, J., Huang, T., Hu, S., Gao, W., Xu, Z., & Geng, J. (2026). Hierarchical memory-enhanced networks for student knowledge tracing. Pattern Recognition Letters. doi:10.1016/j.patrec.2026.01.002
- [14]. Wang, H., & Ji, W. (2026). Cognitive schema-driven knowledge tracing based on hierarchical memory networks: Learning transfer modeling under multi-level cognitive architecture. Expert Systems with Applications, 330, 133050.
doi:10.1016/j.eswa.2026.133050
- [15]. Hu, Q., Wen, Q., Wei, P., Zeng, B., Feng, G., & Fang, L. (2026). Dynamic comprehensive difficulty knowledge cells based on KAN network and stable learning for knowledge tracing. Neural Networks, 197, 108514.
doi:10.1016/j.neunet.2025.108514
- [16]. Tang, Z., He, Z., Ye, G., Li, C., Cao, J., Zhang, Y., & Jiang, J. (2026). UDCL4KT: Uncertainty-aware dual contrastive learning for knowledge tracing. Knowledge-Based Systems, 352, 116902.
doi:10.1016/j.knosys.2026.116902
- [17]. Wu, X., & Wang, Y. (2026). Adaptive learning path generation in virtual reality music education via deep reinforcement learning. Scientific Reports. doi:10.1038/s41598-026-63194-8
- [18]. Chen, X. (2026). An algorithm for dynamic adjustment of personalized education management strategies driven by reinforcement learning. Discover Artificial Intelligence, 6, 368. doi:10.1007/s44163-026-01019-3
- [19]. Fu, Z. (2025). Integrating reinforcement learning with dynamic knowledge tracing for personalized learning path optimization. *Scientific Reports, 15*, 40202.

- <https://doi.org/10.1038/s41598-025-23900-4>
- [20]. Hariyanto, Kristianingsih, F. X. D., & Maharani, R. (2025). Artificial intelligence in adaptive education: A systematic review of techniques for personalized learning. **Discover Education*, 4*, 458. <https://doi.org/10.1007/s44217-025-00908-6>
- [21]. Riedmann, A., Schaper, P., & Lugin, B. (2025). Reinforcement learning in education: A systematic literature review. **International Journal of Artificial Intelligence in Education*, 35*, 2669–2723. <https://doi.org/10.1007/s40593-025-00494-6>
- [22]. Ning, Q., Guo, C., Liu, K., Tang, J., Zhang, S., Zeng, W., & Zhao, X. (2025). Dual sequence modeling for knowledge tracing. *Data Science and Engineering*, 10, 653–664. doi:10.1007/s41019-025-00294-x
- [23]. Wang, S., Zhang, X., Yang, X., Zhang, X., & Wang, K. (2025). Learning states enhanced knowledge tracing: Simulating the diversity in real-world learning process. *Expert Systems with Applications*, 274, 126838. doi:10.1016/j.eswa.2025.126838
- [24]. Luo, H., Zhang, Z., Cui, L., Hu, H., & Yu, B. (2025). Cluster-driven knowledge tracing: Joint learning-forgetting effects modeling via state dependency. *Expert Systems with Applications*, 292, 128619. doi:10.1016/j.eswa.2025.128619
- [25]. Li, Q., Yuan, X., Yue, J., Shen, X., Liang, R., Liu, S., & Yan, Z. (2025). Dual-view multi-scale cognitive representation for deep knowledge tracing. *Knowledge-Based Systems*, 310, 113010. doi:10.1016/j.knosys.2025.113010
- [26]. Sun, X., Liu, Q., Zhang, K., Shen, S., Yang, L., & Li, H. (2025). Harnessing code domain insights: Enhancing programming knowledge tracing with large language models. *Knowledge-Based Systems*, 317, 113396. doi:10.1016/j.knosys.2025.113396
- [27]. Wu, Q., Ji, W., Zhou, G., & Yang, Y. (2025). Graph knowledge tracing in cognitive situation: Validation of classic assertions in cognitive psychology. *Knowledge-Based Systems*, 315, 113281. doi:10.1016/j.knosys.2025.113281
- [28]. Shen, X., Yu, F., Liu, Y., Liang, R., Wan, Q., Yang, T., Shi, M., & Sun, J. (2025). Enhancing knowledge tracing with question-based contrastive learning. *Knowledge-Based Systems*, 325, 113899. doi:10.1016/j.knosys.2025.113899
- [29]. Wen, Y.-F., Liang, H., Yang, C., Zhou, T., Liu, J., Du, Y., & Lee, Y.-L. (2025). Sequential contrastive learning for progressive knowledge tracing. *Knowledge-Based Systems*, 329, 114413. doi:10.1016/j.knosys.2025.114413
- [30]. Wang, J., Ma, H., Zhang, M., Zhang, L., & Chang, L. (2025). Multi-granularity ensemble interaction graph modeling for knowledge tracing. *Knowledge-Based Systems*, 309, 112834. doi:10.1016/j.knosys.2024.112834
- [31]. Yang, H., Hu, J., Chen, J., Hu, S., Geng, J., Zhu, Q., & Huang, T. (2025). MAHKT: Knowledge tracing with multi-association heterogeneous graph embedding based on knowledge transfer. *Knowledge-Based Systems*, 310, 112958. doi:10.1016/j.knosys.2025.112958
- [32]. Jiang, X., Yang, B., & Liu, W. (2025). DMAM: Difficulty-enhanced multi-view attention-based model for knowledge

- tracing. *Knowledge-Based Systems*, 324, 113736.
- [33]. Deng, Y., Guan, Z., He, M., Wang, X., Liu, J., Li, Z., & Yang, Z. (2025). Cross-discipline cold-start knowledge tracing via overlapping entities and contrastive learning. *Knowledge-Based Systems*, 330, 114616. doi:10.1016/j.knosys.2025.114616
- [34]. Yang, L., Sun, X., Li, H., Xu, R., & Wei, X. (2025). Difficulty aware programming knowledge tracing via large language models. *Scientific Reports*, 15, 11475. doi:10.1038/s41598-025-96540-3
- [35]. Zhang, Y., Cao, X., Li, W., & Zhang, L. (2025). A fine-grained course session recommendation method based on knowledge point pruning. *Scientific Reports*, 15, 13164. doi:10.1038/s41598-025-95395-y
- [36]. Hua, Z., Yang, J., & Ji, W. (2025). Knowledge graph convolutional networks with user preferences for course recommendation. *Scientific Reports*, 15, 30256. doi:10.1038/s41598-025-14150-5
- [37]. Liu, H. (2025). Personalized learning support system for special education: A real-time feedback mechanism based on deep reinforcement learning. *Frontiers in Psychology*, 16, 1658698. doi:10.3389/fpsyg.2025.1658698
- [38]. Wang, H., Wu, Q., Bao, C., Ji, W., & Zhou, G. (2025). Research on knowledge tracing based on learner fatigue state. *Complex & Intelligent Systems*, 11, 226. doi:10.1007/s40747-025-01831-x
- [39]. Yang, Q., Chi, J., Chen, W., Wu, Z., Huang, Y., & Zhang, J. (2025). Learning intention-aware knowledge tracing for learning stage. *Discover Computing*, 28, 98. doi:10.1007/s10791-025-09616-z
- [40]. Huo, Y., He, M., Tan, X., & Chen, K. (2025). GCN and GAT-based interpretable knowledge tracing model. *Complex & Intelligent Systems*, 11, 300. doi:10.1007/s40747-025-01921-w
- [41]. Zhang, J. (2025). Graph-based knowledge tracing for personalized MOOC path recommendation. *Journal of Advanced Computing Systems*, 5(11), 1–15. doi:10.69987/JACS.2025.51101
- [42]. Luo, H., Zhang, Z., Cui, L., Zhang, Z., & Liang, Y. (2024). An efficient state-aware coarse-fine-grained model for knowledge tracing. *Knowledge-Based Systems*, 302*, 112375. <https://doi.org/10.1016/j.knosys.2024.112375>
- [43]. Qian, L., Zheng, K., Wang, L., & Li, S. (2024). Student state-aware knowledge tracing based on attention mechanism: A cognitive theory view. *Pattern Recognition Letters*, 184*, 190–196. <https://doi.org/10.1016/j.patrec.2024.06.009>
- [44]. Huang, C., Wei, H., Huang, Q., Jiang, F., Han, Z., & Huang, X. (2024). Learning consistent representations with temporal and causal enhancement for knowledge tracing. *Expert Systems with Applications*, 245, 123128. doi:10.1016/j.eswa.2023.123128
- [45]. Sun, X., Zhang, K., Shen, S., Wang, F., Guo, Y., & Liu, Q. (2024). Target hierarchy-guided knowledge tracing: Fine-grained knowledge state modeling. *Expert Systems with Applications*, 251, 123898. doi:10.1016/j.eswa.2024.123898
- [46]. Huang, T., Ou, X., Yang, H., Hu, S., Geng, J., Xu, Z., & Yang, Z. (2024). Pull together: Option-weighting-enhanced mixture-of-experts knowledge tracing. *Expert Systems*

- p>with Applications, 248, 123419.
-
- doi:10.1016/j.eswa.2024.123419
- [47]. Cui, C., Ma, H., Dong, X., Zhang, C., Zhang, C., Yao, Y., Chen, M., & Ma, Y. (2024). Model-agnostic counterfactual reasoning for identifying and mitigating answer bias in knowledge tracing. *Neural Networks*, 178, 106495. doi:10.1016/j.neunet.2024.106495
- [48]. Park, S., Lee, D., & Park, H. (2024). Enhancing knowledge tracing with concept map and response disentanglement. *Knowledge-Based Systems*, 302, 112346. doi:10.1016/j.knosys.2024.112346
- [49]. Zhang, S., Pu, J., Cui, J., Shen, S., Chen, W., Hu, K., & Chen, E. (2024). MLC-DKT: A multi-layer context-aware deep knowledge tracing model. *Knowledge-Based Systems*, 303, 112384. doi:10.1016/j.knosys.2024.112384
- [50]. Ke, F., Wang, W., Tan, W., Du, L., Jin, Y., Huang, Y., & Yin, H. (2024). HiTSKT: A hierarchical transformer model for session-aware knowledge tracing. *Knowledge-Based Systems*, 284, 111300. doi:10.1016/j.knosys.2023.111300
- [51]. Zhang, K., Ji, T., & Zhang, H. (2024). Knowledge tracing via multiple-state diffusion representation. *Expert Systems with Applications*, 255, 124797. doi:10.1016/j.eswa.2024.124797
- [52]. Huang, T., Hu, S., Yang, H., Geng, J., Li, Z., Xu, Z., & Ou, X. (2024). Response speed enhanced fine-grained knowledge tracing: A multi-task learning perspective. *Expert Systems with Applications*, 238, 122107. doi:10.1016/j.eswa.2023.122107
- [53]. Yang, H., Hu, S., Geng, J., Huang, T., Hu, J., Zhang, H., & Zhu, Q. (2024). Heterogeneous graph-based knowledge tracing with spatiotemporal evolution. *Expert Systems with Applications*, 238, 122249.
- [54]. Zhang, J., Xia, R., Miao, Q., & Wang, Q. (2024). Explore Bayesian analysis in cognitive-aware key-value memory networks for knowledge tracing in online learning. *Expert Systems with Applications*, 257, 124933. doi:10.1016/j.eswa.2024.124933
- [55]. Zhu, M., Qiu, L., & Zhou, J. (2024). Meta-path structured graph pre-training for improving knowledge tracing in intelligent tutoring. *Expert Systems with Applications*, 254, 124451. doi:10.1016/j.eswa.2024.124451
- [56]. Duan, Z., Dong, X., Gu, H., Wu, X., Li, Z., & Zhou, D. (2024). Towards more accurate and interpretable model: Fusing multiple knowledge relations into deep knowledge tracing. *Expert Systems with Applications*, 243, 122573. doi:10.1016/j.eswa.2023.122573
- [57]. Yang, Z., Hu, J., Zhong, S., Yang, L., & colleagues. (2024). Graph-based effective knowledge tracing via subject knowledge mapping. *Education and Information Technologies*. doi:10.1007/s10639-024-13069-0
- [58]. Yang, Y., Chen, H., Liu, X., & Huang, P. (2024). Guidance-as-progressive in human skill training based on deep reinforcement learning. *Journal of Intelligent & Robotic Systems*, 110, 116. doi:10.1007/s10846-024-02147-7
- [59]. Wu, T., & Ling, Q. (2023). Fusing hybrid attentive network with self-supervised dual-channel heterogeneous graph for knowledge tracing. *Expert Systems with Applications*,

- 225, 120212.
doi:10.1016/j.eswa.2023.120212
- [60]. Ni, Q., Wei, T., Zhao, J., He, L., & Zheng, C. (2023). HHSKT: A learner-question interactions based heterogeneous graph neural network model for knowledge tracing. *Expert Systems with Applications*, 215, 119334. doi:10.1016/j.eswa.2022.119334
- [61]. Zhao, Y., Ma, H., Wang, W., Gao, W., Yang, F., & He, X. (2023). Exploiting multiple question factors for knowledge tracing. *Expert Systems with Applications*, 223, 119786. doi:10.1016/j.eswa.2023.119786
- [62]. Abdelrahman, G., & Wang, Q. (2023). Learning data teaching strategies via knowledge tracing. *Knowledge-Based Systems*, 269, 110511. doi:10.1016/j.knosys.2023.110511
- [63]. Kim, J., Koo, S., & Lim, H. (2023). A multi-faceted exploration incorporating question difficulty in knowledge tracing for English proficiency assessment. *Electronics*, 12(19), 4171. doi:10.3390/electronics12194171
- [64]. Chen, X., Shen, J., Xia, W., Jin, J., Song, Y., Zhang, W., Liu, W., Zhu, M., Tang, R., Dong, K., Xia, D., & Yu, Y. (2023). Set-to-sequence ranking-based concept-aware learning path recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(4), 5027–5035. doi:10.1609/aaai.v37i4.25630
- [65]. Ma, Y., Wang, L., Zhang, J., Liu, F., & Jiang, Q. (2023). A personalized learning path recommendation method incorporating multi-algorithm. *Applied Sciences*, 13(10), 5946. doi:10.3390/app13105946
- [66]. Li, T., Wang, X., Zhang, S., Yang, F., & Lu, W. (2023). A personalized learning path recommendation method for learning objects with diverse coverage levels. In *Artificial Intelligence in Education* (pp. 714–719). Springer. doi:10.1007/978-3-031-36272-9_61
- [67]. Zhang, X., Liu, S., & Wang, H. (2023). Personalized learning path recommendation for e-learning based on knowledge graph and graph convolutional network. *International Journal of Software Engineering and Knowledge Engineering*, 33(1), 109–132. doi:10.1142/S0218194022500681
- [68]. Aberbach, H., Sabri, A., & Marappan, R. (2023). An automatic recommendation process to generate learning paths based on learner preferences. *International Journal of Information and Education Technology*, 13(10), 1549–1555. doi:10.18178/ijiet.2023.13.10.1961.