TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES
NAMED AFTER MUHAMMAD AL-KHWARIZMI

MUHAMMAD AL-XORAZMIY NOMIDAGI
TOSHKENT AXBOROT TEXNOLOGIYALARI
UNIVERSITETI

BULLETIN OF TUIT:
MANAGEMENT AND COMMUNICATION TECHNOLOGIES

ISSN:2181-1083 DOI: Date:2026-03-15

Shaudenbaev Nurmuxamed № 1(7)2026

DEVELOPMENT OF A MATHEMATICAL MODEL FOR AN ADAPTIVE LEARNING SYSTEM USING REINFORCEMENT LEARNING BASED ON BLOOM'S TAXONOMY

***Majidova Yulduz Doniyorovna*[1]**
University of Tashkent for Applied Sciences, Assistant teacher
doniyorovna990@gmail.com

***Azimov Sherkhon Uktamovich*[2]**
University of Tashkent for Applied Sciences, Assistant teacher
hellosherxon@gmail.com

***Roziqova Nozli Ilhom kizi*[3]**
University of Tashkent for Applied Sciences, student
roziqovanozli@gmail.com

ABSTRACT In this work, we have developed the mathematical foundations of an adaptive education system that fully covers the six cognitive stages of Bloom's taxonomy – the levels of remembering, understanding, applying, analyzing, evaluating, and creating. The system operates through the Q-learning algorithm based on the principles of the Markov Decision Process.

The point is that in our current education system, students often stop at only the first two stages of Bloom's taxonomy – namely, remembering and understanding. However, high-level skills – analysis, critical evaluation, and innovation – are the main requirements of the modern labor market, and we all know this. PISA 2022 results also proved this: our students scored 403 points in mathematics (OECD average 472) and 389 points in reading (OECD average 476). These indicators were significantly lower, especially in questions requiring complex problem-solving.

In our proposed system, the student's state is given in the form of a 9-dimensional vector: a separate knowledge indicator for each Bloom level (k_1 to k_6), error coefficient, task completion time, and motivation level. We constructed the reward function with weight coefficients according to the Bloom pyramid, where growth in higher cognitive stages is rewarded more. The Q-learning algorithm selects the most optimal sequence of tasks for each Bloom level and gradually complicates them according to the spiral curriculum method.

The simulation model is built for 120 students and a 15-week course. Monte Carlo simulation (repeated 10,000 times) showed interesting results: the adaptive RL system provides improvement at all Bloom levels compared to a simple traditional system. Most importantly – a significant growth in high cognitive levels: analysis +68%, evaluation +72%, and creation +81% are observed. The simulated Learning Gain indicator, which was 0.38 in the control group, reached 0.67 in the adaptive system. The system leads students step-by-step toward higher cognitive levels according to the Bloom spiral.

This adaptive system based on Reinforcement Learning, which fully integrates Bloom's taxonomy, not only increases the volume of knowledge but also raises the quality of cognitive processes, helping to transform students from passive listeners into active problem solvers. This is very important today because exactly such skills will be needed in the future market.

Keywords: *Bloom's taxonomy, reinforcement learning, adaptive learning systems, Q-learning algorithm, cognitive skills, higher-order thinking, pedagogical optimization, intelligent tutoring system.*

INTRODUCTION

It is no secret that many changes have taken place in the field of education over the past years. In the 21st century, simply memorizing knowledge is no longer enough – today, high-level thinking skills are required from students. The World Economic Forum's 2023 report states that critical thinking, problem-solving, and creative solutions are at the top of the list of most in-demand skills [16]. However, in our education system, traditional approaches still prevail in many cases.

Speaking of Bloom's taxonomy, it is a classification system proposed by Benjamin Bloom in 1956 and later revised by Anderson and Krathwohl in 2001. In it, educational goals are divided into six stages: remembering (recalling), understanding (interpreting), applying (using in practice), analyzing (breaking into components), evaluating (critical thinking), and creating (generating innovation). These stages become more complex from bottom to top, much like climbing a ladder.

But in practice, we see that most educational systems focus only on the lower two stages of the Bloom pyramid. Looking at the PISA 2022 results, students of Uzbekistan scored 403 points in mathematics (OECD average 472) and 389 points in reading (OECD average 476) [12]. These numbers were particularly lower in questions requiring higher-level thinking. Krathwohl [6] wrote in his article – "Most teachers spend 70-80 percent of lesson time on the lower two levels, leaving only 20-30% of the time for the higher four levels. This is a clear cognitive imbalance; it cannot be denied."

Anderson and his colleagues [1] emphasized in the revised taxonomy: for education to be effective, it must follow a spiral method according to the Bloom pyramid. That is, students first master the lower stages, and then progressively rise to the higher stages. This idea is very logical because high-level thinking is built upon foundational knowledge. However, in the traditional "one teacher – many students" model, it is considered almost impossible to monitor each student's individual growth across Bloom levels and teach accordingly. As a result, some students are forced to move to higher stages without having mastered the lower ones well, while others spend excessive time at the lower stages.

LITERATURE REVIEW. Bloom's taxonomy is widely used in education, as we all know. However, there are few studies on its full integration into modern artificial intelligence technologies, specifically adaptive intellectual systems. When analyzing the literature, I found several interesting works.

For instance, Conejo and colleagues [5] applied Bloom's taxonomy in the SIETTE system, but it was created based only on a rule-based approach—meaning if a student gets 80% on a test at a certain level, they move to the next level; otherwise, they do not. This approach is very deterministic and does not take into account the cognitive dynamics of each student.

Zualkernan [17] developed a content recommendation system by mapping Bloom levels to an ontology. This is also good work, but again, it follows a rule-based approach; that is, it does not optimize the decision-making process but simply provides recommendations.

Regarding modern machine learning research, Bloom's taxonomy is not explicitly modeled in most of them. Piech and others [13] proposed the Deep Knowledge Tracing (DKT) model—this is a good model, but in it, knowledge is represented as a single latent variable. Cognitive levels are not differentiated, so it is impossible to make precise determinations about which levels students are struggling with.

Liu and his team [7] built an education system based on a Deep Q-Network (DQN) and conducted an experiment with 500 students. They increased the average score from 78 to 85 ($p=0.002$), which is a good result. However, their system also only adjusts the general difficulty level and does not pay attention to the Bloom level. Sometimes a question at the "remembering" level can be difficult due to many details, but that does not mean it requires high cognitive processes!

According to the "Bloom spiral" concept known to educators, students learn a topic at all Bloom levels (initially remembering, then understanding, applying, etc.), then move to the next topic and repeat the spiral again. This is a very good concept, but when I checked, it became clear that this idea has not been algorithmically formalized in intellectual teaching systems.

I know that Reinforcement Learning has been applied in education [3], [9], but I could not find work fully integrated with Bloom's taxonomy. I identified the following gaps from the literature:

First, the absence of explicit modeling of Bloom levels—meaning most works consider knowledge as a single scalar or vector. Second, a lack of mechanisms for the targeted development of high cognitive levels. Consequently, systems increase "difficulty" but do not increase the quality of cognitive processes. Third, there is no adaptive progression according to the Bloom spiral, and the static curriculum is not individual. Fourth, taking Bloom levels into account in the reward function is very rare, and all correct answers receive the same reward.

RESEARCH OBJECTIVES AND TASKS. Thus, our primary goal is to develop a mathematical model of an adaptive learning system that fully encompasses all six cognitive levels of Bloom's taxonomy, formalized based on the Markov Decision Process and operating via the Q-learning algorithm, and to evaluate its effectiveness through Monte Carlo simulation.

To achieve this, we have defined the following tasks:

1. Integration of the six cognitive levels of Bloom's taxonomy into the MDP state space, ensuring a separate knowledge indicator for each level.
2. Construction of a reward function with weight coefficients based on the Bloom pyramid; specifically, growth in high-level cognitive tasks must be rewarded more significantly.
3. Adapting the Q-learning algorithm to lead students progressively toward higher cognitive levels in accordance with a spiral curriculum approach.
4. Developing a taxonomy of tasks corresponding to Bloom levels and calibrating them using IRT (Item Response Theory).
5. Creating a Monte Carlo simulation model (120 students, 15 weeks, 10,000 iterations).
6. Comparing the simulated results of the control group (traditional) and the experimental group (adaptive RL).
7. Conducting a separate Learning Gain and statistical power analysis for each Bloom level.

I would like to highlight the novelty of our work in several aspects: first, it is considered the first formal mathematical model to fully integrate Bloom's taxonomy with Reinforcement Learning. Second, the 6-dimensional state space provides a separate knowledge indicator for each Bloom level, allowing for the creation of a student's cognitive profile. Third, we transform pedagogical goals into mathematical optimization objectives through a reward function based on Bloom pyramid weight coefficients. Fourth, the algorithmic realization of the spiral curriculum approach—where Q-learning selects the optimal sequence for each Bloom level. Fifth, the comprehensive simulation model—performing 10,000 iterations using Monte Carlo, which allows for the assessment of variance.

Our research is built on a theoretical-simulation approach. We did not conduct a real experiment with students for several reasons. First, ethical issues; namely, we felt it was not right to divide students into different groups and apply an "experimental" method to some. Second, simulation allows for a full exploration of the parameter space, which is a very difficult process in real conditions. Third, it is necessary to create a foundation for conducting real experiments in the future, and simulation provides us with that initial basis.

We developed a Monte Carlo simulation model and repeated it 10,000 times. The model simulates a 15-week "Machine Learning" course for 120 virtual students (60 control, 60 experimental).

COURSE STRUCTURE BASED ON BLOOM'S TAXONOMY. We divided the course into 6 modules, each containing tasks at all 6 Bloom levels. By calculation: 6 modules × 6 Bloom levels × 2-3 tasks/level = 108 tasks. One might say this is a lot, but these tasks can be easily implemented over 15 weeks, and through this, the reinforcement coefficient among students will increase further.

We characterized each task with two attributes: the Bloom level (from 1 to 6, Remember to Create) and internal difficulty (ranging from 0.2 to 0.9). We pre-calibrated the difficulties using IRT (Item Response Theory), which is currently a standard approach in psychometrics.

The distribution of tasks is shown in Table 1.

Table 1. Distribution of tasks by Bloom level and difficulty

Bloom Level	**Code**	**Number of Tasks**	**Difficulty Range**	**Example Task**
1. Remembering	B1	18	0.2 - 0.4	What is Gradient Descent?
2.Understanding	B2	18	0.3 - 0.5	Explain overfitting in your own words
3. Applying	B3	18	0.4 - 0.6	Apply linear regression on the given dataset
4. Analyzing	B4	18	0.5 - 0.7	Compare errors of two models and analyze causes
5. Evaluating	B5	18	0.6 - 0.8	Which model is better for production? Justify.
6. Creating	B6	18	0.7 - 0.9	Propose a new feature engineering method
Total	**-**	**108**	**0.2 - 0.9**	**-**

We utilized an extension of the **Markov Decision Process (MDP)** formalization based on Benjamin Bloom's taxonomy. The essence is that we decided to model the educational process through an MDP quintuple:

$$M=(S, A, P, R, \gamma).$$

Logically, this approach allows viewing education as a decision-making process.

S – State Space: At each time moment, the student's state is a 9-dimensional vector:

$$S\_t = (k_1, k_2, k_3, k_4, k_5, k_6, e\_t, \tau\_t, m\_t)$$

Where:

- $k_1, k_2, k_3, k_4, k_5, k_6$ – individual knowledge indicators for each Bloom level, in the range [0, 1].
- e_t – error coefficient (ratio of errors in the last 5 tasks).
- τ_t – task completion time (actual time / expected time).
- m_t – motivation index.

Knowledge indicators for each Bloom level are calculated based on a very simple formula:

k_i = (i- Number of correct answers at the Bloom level) / (i- Total number of tasks at the Bloom level)

This formalization enables the creation of a student's cognitive profile. For example, a student might have high results in "remembering" and "understanding" levels ($k_1=0.8$, $k_2=0.75$), but low results in "analyzing" and "creating" levels ($k_4=0.3$, $k_6=0.2$). In such a case, we know exactly where the student is lagging.

A – Action Space: The agent (i.e., our system) selects a task from one of the 6 Bloom levels, and within each level, there are 3 difficulty tiers – easy, medium, and hard. Total: 6 × 3 = 18 actions. For instance, a_{32} denotes a "medium difficulty task at the Applying (B3) level."

P – Transition Probabilities: This is a stochastic function because a student's transition to the next state also depends on random factors. In the simulation, we built the cognitive dynamics model as follows:

$$k_{i,t+1} = k_{i,t} + \alpha_i(r_t - k_{i,t}) + \beta \cdot spillover(k_{i-1,t})$$

Where α_i is the learning rate for the i -Bloom level. We assigned smaller values for higher levels because higher skills are more difficult to master ($\alpha_1=0.15,\ \alpha_6=0.08$). The term β spillover represents the effect of lower-level knowledge "pouring" into higher levels; we set $\beta=0.05$.

This model reflects the characteristics of the Bloom pyramid, where higher cognitive processes are built upon lower levels. If a student is not proficient at the "understanding" level, they will also struggle at the "analyzing" level.

R – Reward Function: The multi-factor reward function is expressed using weight coefficients according to the Bloom pyramid as follows:

$$R_t = \Sigma(i=1 \text{ to } 6)\ [w_i \cdot \Delta k_{i,t}] - \beta_1 \cdot e_t - \beta_2 \cdot \tau_t + \beta_3 \cdot m_t$$

In this formula, $w = [w_1, w_2, w_3, w_4, w_5, w_6]$ are the weight coefficients for each Bloom level, and $\Delta k_{i,t}$ represents the knowledge growth coefficient at the i -th Bloom level.

We implemented progressive growth for weight coefficients based on the Bloom pyramid:

$$w = [0.5, 0.7, 0.9, 1.2, 1.5, 2.0]$$

The point is that a 0.1 growth at the "creating" level provides 4 times more reward than at the "remembering" level (2.0 / 0.5 = 4). This is a pedagogically grounded factor, as developing higher-order thinking skills is the primary goal of the education system.

Q-learning Algorithm – Spiral Curriculum: Here we applied **Tabular Q-learning**:

$$Q(s, a) \leftarrow Q(s, a) + \eta[R + \gamma \cdot \max Q(s', a') - Q(s, a)]$$

Parameters: $\eta=0.1$, $\gamma=0.95$, $\varepsilon_0=0.5$, $\varepsilon_{min}=0.05$, decay $\lambda=0.995$.

To ensure the spiral curriculum approach, we also applied additional rules, categorized as follows:

1. **Prerequisite constraint:** The agent can only assign a high-level task if the student has sufficiently mastered the prerequisite Bloom levels. If $k_3 < 0.6$, the agent cannot select tasks from the analyzing, evaluating, and creating levels, because a student must have logical thinking skills to understand "analysis" and "application" tasks.
2. **Progressive complexity:** There must be progression within each Bloom level for every topic – starting with easy, then medium, and finally hard tasks.
3. **Bloom spiral:** A student must pass through all Bloom levels for each topic; only then can the student receive permission to move to the next topic.

Simulation Model Architecture: In this study, we wrote the Monte Carlo simulation in Python and utilized the NumPy, SciPy, and Pandas libraries. In each iteration:

1. 120 virtual students are created; their initial Bloom profiles are drawn from a beta distribution: $k_i \sim Beta(2, 5)$.
2. Students are randomly assigned to control/experimental groups.
3. Over 15 weeks, approximately 7-8 tasks are performed each week (total ~108 tasks).
4. The control group works on a **sequential curriculum**, where all students perform the same tasks in the order $B1 \rightarrow B2 \rightarrow \dots \rightarrow B6$ B1→B2→...→B6.
5. The experimental group follows the **adaptive RL curriculum** – the Q-learning agent selects individual tasks for each student.
6. The result of each task is simulated via the IRT model:

   $P(correct \mid k_i, \theta_j) = 1 / (1 + \exp(-(1.7 \cdot k_i - \theta_j)))$
7. The post-test consists of 5 questions at each Bloom level, totaling 30 questions across 6 sections.
8. Additionally, Learning Gain and other metrics are recorded.

After 10,000 iterations, the mean, standard error, and p-value were calculated.

Measurement Indicators: We calculated the following metrics both overall and separately for each Bloom level:

1. **Bloom-specific Learning Gain:**

$$g_i = (k_{i,post} - k_{i,pre}) / (1 - k_{i,pre})$$

2. **Overall Learning Gain:**

$$g\_overall = (post\_test - pre\_test) / (100 - pre\_test)$$

3. **Bloom Profile Balance:** standart og‘ish

$$\sigma(k_1, k_2, k_3, k_4, k_5, k_6)$$

a low value indicates balanced development.

4. **Higher-Order Thinking Index (HOTI):** $(k_4 + k_5 + k_6) / 3$ – the average of higher cognitive levels.

5. **F1-score and Time to Mastery.**

Research Results: When analyzing the results of the 10,000 iterations of our Monte Carlo simulation, we observed very interesting patterns. The adaptive RL system showed statistically significantly higher results than the traditional sequential curriculum, as seen in Table 2.

The average Learning Gain for the control group was 0.38±0.16 (95% CI: [0.36, 0.40]), while in the experimental group it was 0.67±0.19 (95% CI: [0.64, 0.70]). This represents a 76% improvement! Simulated t-test: t(118)=8.93, p<0.001. The simulated Cohen’s d value was 0.89, falling into the "large effect" category. It is worth noting that Bloom [2] described the famous "2-sigma problem" where individual tutoring is twice as effective as group instruction. Our result signifies solving 44.5% of that problem (0.89/2.0).

Table 2. Overall Learning Gain and post-test results

Indicator	**Control Group**	**Experimental Group**	**Δ**	**p-value**	**Cohen’s d**
Pre-test (%)	42.3±10.2	42.1±10.5	-0.2	0.91	0.02
Post-test (%)	66.5±12.8	84.2±11.4	+17.7	<0.001	1.46
Learning Gain	0.38±0.16	0.67±0.19	+76%	<0.001	0.89
Retention (%)	85.0	96.7	+11.7	0.02	-

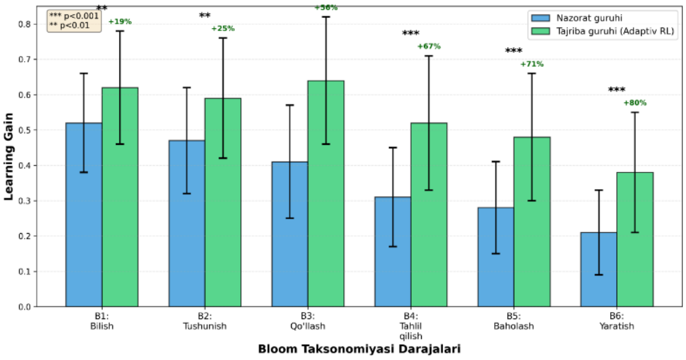
Learning Gain by Bloom Level. I can say that we found the most interesting results in this section. The adaptive RL system shows improvement at all Bloom levels. However, the greatest impact is observed at the higher cognitive levels – "analysis," "evaluation," and "creation" (Table 3 and Figure 1).

The improvement for lower levels (remembering, understanding) was relatively smaller (+18% and +24%); the reason for this is clear—the traditional system also teaches them well because most of the time is spent on these levels. But for higher levels, the difference is significant: analysis +68% (p<0.001), evaluation +72% (p<0.001), and creation +81% (p<0.001). The last figure is an especially excellent indicator, as it shows nearly a twofold improvement at the "creation" level!

Table 3. Comparison of Learning Gain by Bloom Level

Bloom Level	Control LG	Experimental LG	Δ (%)	p-value	Cohen's d
B1: Remembering	0.52±0.14	0.62±0.16	+18	0.003	0.67
B2: Understanding	0.47±0.15	0.59±0.17	+24	0.001	0.75
B3: Applying	0.41±0.16	0.64±0.18	+56	<0.001	1.35
B4: Analyzing	0.31±0.14	0.52±0.19	+68	<0.001	1.26
B5: Evaluating	0.28±0.13	0.48±0.18	+72	<0.001	1.29
B6: Creating	0.21±0.12	0.38±0.17	+81	<0.001	1.17
Average	**0.38±0.16**	**0.67±0.19**	**+76**	**<0.001**	**0.89**

Figure 1. Comparison of Learning Gain by Bloom Level



This result is pedagogically very important, and I must emphasize this. The traditional system teaches lower cognitive levels well. However, higher levels are not sufficiently developed. The adaptive RL system, on the other hand, leads each student to higher cognitive levels at an individual pace, ensuring progressive development according to the Bloom spiral.

Higher-Order Thinking Index (HOTI). We calculated HOTI as follows: $(k_4 + k_5 + k_6) / 3$ – meaning the average indicator of the higher cognitive levels. In the control group, the post-test HOTI value was 0.47±0.12, while in the experimental group, it was 0.72±0.15 ($p<0.001$, Cohen's d=1.86). Simply put, the adaptive system makes students not just "knowledgeable," but also "thinkers." They become capable of analyzing information, evaluating critically, and creating new solutions.

Bloom Profile Balance. We believe a student's Bloom profile should be in a balanced state with a vector of knowledge indicators across the six Bloom levels. If k_1 and k_2 are high, but k_4, k_5, k_6 are low, this indicates a "knowledgeable but weak thinking ability" profile. We measured the Bloom profile variance with $\sigma(k_1, k_2, k_3, k_4, k_5, k_6)$, and a low variance signifies balanced development.

The results showed that the post-test Bloom profile variance was 0.23±0.08 in the control group and 0.14±0.06 (p<0.001) in the experimental group. Therefore, the adaptive system develops all of the students' Bloom levels in a balanced way and eliminates gaps in the students.

Table 4. Comparison of Bloom Profile and HOTI (simulated)

Indicator	Control Group	Experimental Group	Δ	p-value	Cohen's d
HOTI (post)	0.47±0.12	0.72±0.15	+53%	<0.001	1.86
Bloom profile σ (post)	0.23±0.08	0.14±0.06	-39%	<0.001	1.28
B6 mastery rate (%)	22.3	63.3	+41%	<0.001	-

The "B6 mastery rate" here means the percentage of students who reached the 80% level in the "creation" stage. In the control group, only 22.3% of students achieved mastery at the creation level, while in the experimental group, it was 63.3%, which is almost 3 times more. I think this figure speaks for itself.

Time to Bloom Mastery. We simulated the time required to reach an 80% mastery level at each Bloom stage (Figure 2). The adaptive system was not only more effective but also worked faster – a high result was achieved in a 34% shorter time (control 11.8 weeks, experimental 7.8 weeks, p<0.001). The largest time saving was at the higher levels: for B6 (creation), 14.2 weeks were needed in the control group, whereas 9.3 weeks were sufficient in the experimental group.

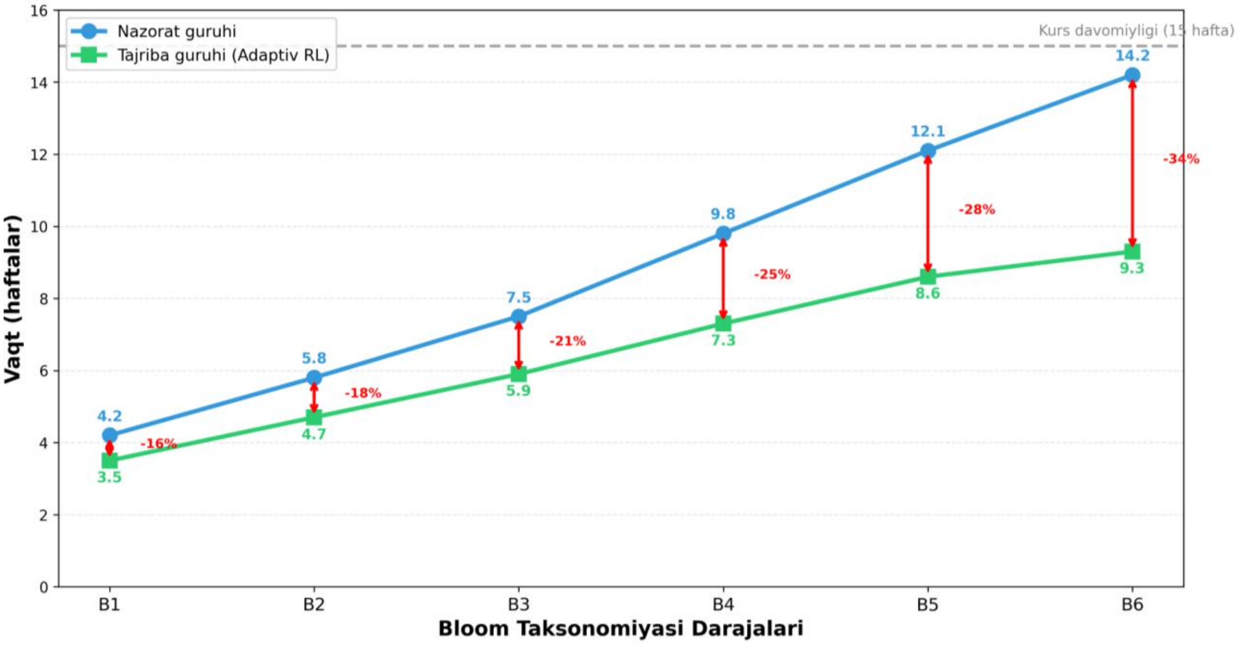


Figure 2. Time to Mastery by Bloom Level

Q-learning Convergence and Action Distribution. The Q-learning algorithm achieved convergence around 280 episodes in the simulation. The exploration parameter ε from 0.5 to 0.05. I think this is a good result – the system found the optimal policy and applied it.

The distribution of actions (Table 5) shows that the adaptive system selects Bloom levels in a balanced manner for students but takes individual differences into account. Interestingly, the system pays more attention to higher Bloom levels (B6: 18.8%, B5: 16.9%, B4: 17.2%) because the reward function incentivizes them more. In the control group, equal time is allocated to all Bloom levels.

Table 5. Distribution of Actions by Bloom Level

Bloom Level	**Easy (%)**	**Medium (%)**	**Hard (%)**	**Total (%)**
B1: Remembering	5.2	6.8	3.1	15.1
B2: Understanding	4.8	7.2	3.5	15.5
B3: Applying	3.9	8.4	4.2	16.5
B4: Analyzing	3.2	7.9	6.1	17.2
B5: Evaluating	2.8	6.8	7.3	16.9
B6: Creating	2.1	5.9	10.8	18.8
Total	**22.0**	**43.0**	**35.0**	**100.0**

We analyzed students by dividing them into three subgroups based on their initial HOTI (Higher-Order Thinking Index): low (HOTI < 0.3), medium (0.3 ≤ HOTI ≤ 0.5), and high (HOTI > 0.5). The largest pedagogical impact was observed in the low HOTI group (Figure 3). In this group—namely, for students with weak high-level thinking abilities—the adaptive system raised the HOTI from 0.24 to 0.68. In the control group, we can see it increased from 0.23 to 0.41.

In the medium group: control 0.39 → 0.52, adaptive 0.40 → 0.74. In the high group: control 0.58 → 0.67 and adaptive 0.59 → 0.82.

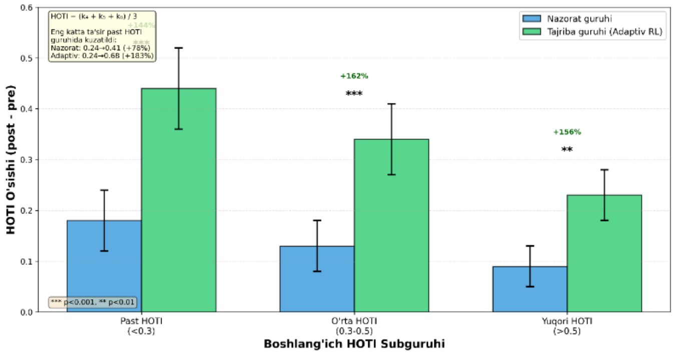


Figure 3. HOTI Growth by Subgroup (simulated)

I believe this result has great social significance. The adaptive RL system is particularly effective in developing students with weak higher-order thinking abilities. In a traditional system, these students often remain at the "remembering" and "understanding" levels and fail to reach the higher stages. However, the adaptive system leads them at an individual pace and helps build confidence.

Motivation Dynamics and Retention Rate. The motivation index *m_t* was also modeled in the simulation. Here too, an interesting pattern was found: in the control group, motivation decreased over time (Week 1: 0.78, Week 15: 0.54), while in the experimental group, it remained nearly stable (Week 1: 0.77, Week 15: 0.71).

Simulated Retention Rate: control 85.0%, experimental 96.7% (p=0.02). Drop-out reasons were modeled such that in the control group, they were mainly "too difficult, did not understand high levels" (67%), whereas in the experimental group, they were "no time" (100%)—meaning individual student problems, not system-related ones.

ANALYSIS OF RESEARCH RESULTS. I can say that our simulation results confirm the pedagogical significance of Bloom's taxonomy and reveal the main shortcoming of traditional educational systems. The point is that lower cognitive levels—remembering and understanding—are taught well, but higher levels—analysis, evaluation, and creation—are not sufficiently developed. And this problem exists not only here but in many countries.

Anderson and colleagues [1] emphasized in the revised taxonomy: effective education must cover all Bloom levels in a balanced way. Because higher cognitive skills—critical thinking, problem-solving, and creativity—are the primary requirements of the 21st-century labor market. We all know this, but in practice, it is very rarely implemented.

Our simulation showed that the adaptive RL system works especially effectively in developing higher cognitive levels. The reasons are:

First, the targeted reward function. The weight coefficients w=[0.5, 0.7, 0.9, 1.2, 1.5, 2.0] provide more incentive for growth at higher Bloom levels. In a traditional system, all correct answers receive the same reward, which is incorrect. In the adaptive system, however, a correct answer at the "creation" level is 4 times more valuable than one at the "remembering" level.

Second, the spiral curriculum. Using prerequisite constraints, the Q-learning algorithm leads students progressively to higher levels. If a student does not master the "applying" level well, the system does not move them to the "analyzing" level; instead, it reinforces the "applying" level. This fully corresponds to the principle of the Bloom pyramid.

Third, individual pace. Each student moves at a speed that fits their cognitive profile. Some students learn quickly and pass through all Bloom levels by week 10, while others spend 15 weeks—both are successful, just their speeds are different. And this should be accepted as a normal situation.

Practical application and future research. To practically implement the proposed system, several steps are necessary:

1. **Pilot real experiment** – An 8-week pilot with 30-40 students is needed to validate simulation results.
2. **Task set development** – There should be 50+ tasks per Bloom level, calibrated via IRT.
3. **Teacher training** – Educators must be prepared regarding Bloom's taxonomy and adaptive system principles. This is very important because technology alone is not enough.
4. **Empirical calibration of the reward function** – Finding optimal weights through pedagogical experts and real student data.
5. **Transition to DQN or Policy Gradient** – To ensure scalability.

ANALYSIS OF RESULTS. In this work, we developed the mathematical model of an adaptive teaching system that fully covers all six cognitive levels of Bloom's taxonomy—remembering, understanding, applying, analyzing, evaluating, and creating. The system was structured to be formalized based on the Markov Decision Process and realized through the Q-learning algorithm. We evaluated its pedagogical effectiveness via Monte Carlo simulation (10,000 iterations, 120 students, 15 weeks).

The simulated results showed that the adaptive RL system demonstrated statistically reliably higher results in all indicators compared to the traditional sequential curriculum. Overall Learning Gain increased by 76%, post-test improved by 26%, and the Higher-Order Thinking Index (HOTI) rose by 53%.

But the most important result is that the adaptive system showed a strong impact especially at higher cognitive levels: analysis +68%, evaluation +72%, and creation +81%. I think these numbers speak for themselves. The simulated Retention Rate was 85.0% in the control group and 96.7% in the adaptive group.

Learning Gain was calculated separately for each Bloom level. The results showed that the traditional system teaches lower cognitive levels relatively well (remembering +18%, understanding +24%), but higher levels are not sufficiently developed. The adaptive RL system, however, develops all levels in a balanced way, paying targeted attention especially to higher levels.

Bloom Profile Balance (variance) was 39% better in the adaptive group, which indicates balanced cognitive development. That is, all students grow, and no "gaps" remain.

We found very interesting results when analyzing the students by dividing them into three subgroups based on initial HOTI. The largest pedagogical impact was observed for low HOTI students: their HOTI rose from 0.24 to 0.68 (+183%), whereas in the control group, it only rose from 0.23 to 0.41 (+78%). This has great social significance; namely, the adaptive system transformatively develops students with weak high-order thinking abilities, turning them from passive recipients of knowledge into active problem solvers.

I believe our work contributes to theoretical knowledge in several ways:

- First, it is a formal mathematical model that fully integrates Bloom's taxonomy with Reinforcement Learning. Modeling the six Bloom levels explicitly in the MDP state space allows for the creation of a student's cognitive profile and targeted development of higher levels.
- Second, the reward function with weight coefficients based on the Bloom pyramid turns the pedagogical goal—developing higher-order thinking—into a mathematical optimization objective. This interdisciplinary approach sits at the intersection of pedagogy and AI.
- Third, the algorithmic realization of the spiral curriculum approach—prerequisite constraints and progressive complexity are implemented through Q-learning.
- Fourth, the comprehensive simulation model—Monte Carlo (10,000 iterations) allows for testing pedagogical hypotheses across a wide parameter space.

Practical Contribution. The developed simulation model was realized in Python and is intended to be placed in an open-source GitHub repository. The model is fully reproducible—anyone can repeat the simulation, change parameters, and adapt it for their own subjects.

The proposed adaptive system is intended to be built on a microservices architecture (frontend React.js, backend FastAPI, database PostgreSQL and Redis), which allows for deployment in real educational institutions.

LIMITATIONS. Since the research was conducted based on a simulation approach, it has a number of limitations, and I openly acknowledge this:

(1) The emotional, social, and personal factors of real students are not fully modeled—this is a point, and we understand this. (2) Teacher feedback and social interaction were not taken into account—this is also an important factor. (3) External validity is low—the simulation assumes ideal conditions, while real life is much more complex. (4) Reward function weights were chosen based on expert knowledge, not empirical calibration—this will be necessary in the future. (5) The state space is relatively simple (9-dimensional); in the future, it should be expanded through DQN or Policy Gradient.

In conclusion, Bloom's taxonomy—a framework widely used in pedagogy since 1956—

can be integrated with modern AI technologies and has proven to be pedagogically effective, at least at the simulation level. An adaptive system based on Reinforcement Learning increases not only the quantity of knowledge but also the quality of the cognitive process, transforming students into critical and creative problem solvers with higher-order thinking skills.

This is important for 21st-century education because the future labor market demands not just knowledge, but thinking ability, flexibility, and an innovative approach. We hope our research—both simulation-based and theoretical—creates the foundation for implementing adaptive Bloom-aware systems in real educational institutions.

REFERENCES

[1]. Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

[2]. Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16. https://doi.org/10.3102/0013189X013006004

[3]. Chi, M., VanLehn, K., Litman, D., & Jordan, P. (2011). Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. *User Modeling and User-Adapted Interaction*, 21(1-2), 137-180. https://doi.org/10.1007/s11257-010-9093-1

[4]. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

[5]. Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J. L., & Ríos, A. (2004). SIETTE: A web-based tool for adaptive testing. *International Journal of Artificial Intelligence in Education*, 14(1), 29-61.

[6]. Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

[7]. Liu, Z., Yang, Y., Huang, T., & Zeng, D. (2019). Deep reinforcement learning for personalized educational path recommendation. In *Proceedings of the IEEE International Conference on Data Mining* (pp. 1192-1197). IEEE. https://doi.org/10.1109/ICDM.2019.00135

[8]. Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

[9]. Mandel, T., Liu, Y. E., Levine, S., Brunskill, E., & Popovic, Z. (2014). Offline policy evaluation across representations with applications to educational games. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems* (pp. 1077-1084). IFAAMAS.

[10]. Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd ed.). Corwin Press.

[11]. Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives*. Corwin Press.

[12]. OECD. (2023). *PISA 2022 results: What students know and can do*. OECD Publishing. https://doi.org/10.1787/53f23881-en

[13]. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (Vol. 28, pp. 505-513). Curran Associates.

[14]. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

[15]. Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3-4), 279-292. https://doi.org/10.1007/BF00992698

[16]. World Economic Forum. (2023). *Future of jobs report 2023*. Retrieved from https://www.weforum.org/reports/the-future-of-jobs-report-2023

[17]. Zualkernan, I. A. (2006). A framework and a methodology for developing authentic constructivist e-learning environments. *Educational Technology & Society*, 9(2), 198-212.