UDC 005.471

«ADAPTIVE WEB PLATFORM FOR UNT PREPARATION: PROTOTYPE IMPLEMENTATION AND PILOT TESTING WITH STUDENTS»

Оспан Асанәлі Алмазұлы
asanaliiospan@gmail.com
магистрант 2 курса образовательной программы “Computer Science and Engineering”
Astana IT University, г. Астана, Республика Казахстан
Научный руководитель - PhD, ассистент профессор Мырзакеримова А.

I. Introduction

The Unified National Test (UNT) is a high-stakes examination in Kazakhstan because it influences both secondary school graduation and access to higher education [1]. However, student preparation for the UNT is still largely based on uniform materials, repetitive drills, and generalized practice tests [3]. Such approaches may support repetition, but they do not adequately address differences in students’ knowledge level, pace of learning, or topic-specific weaknesses. As a result, many learners follow the same preparation path even though their actual academic needs differ significantly [2].

Mathematics is especially suitable for adaptive preparation because it includes multiple interconnected topics and a clearly structured exam specification. The UNT mathematics component covers areas such as numbers, equations, inequalities, sequences, mathematical analysis, planimetry, stereometry, and vectors [6]. In addition, the subject includes different item formats and several levels of difficulty, which makes topic-level diagnostics and personalized feedback particularly relevant.

In previous work, an adaptive educational platform for UNT preparation was proposed as a machine-learning-oriented concept aimed at providing personalized support and recommendation logic [7]. That stage mainly focused on theoretical justification and system design, rather than on practical implementation with student usage data. Therefore, the next important step is to move from concept to prototype and evaluate whether such a system can operate in practice.

This paper presents the implementation of a prototype adaptive web platform and its preliminary pilot testing with student data. The prototype follows the broader architecture of the dissertation project and supports topic-based testing, performance tracking, and structured logging of assessment attempts. The current pilot is limited to mathematics, which serves as the first subject for implementation and analysis.

The aim of this study is to evaluate whether student performance indicators can be aggregated at the topic level and used to identify weaker and stronger areas in preparation. The practical contribution of the paper is twofold. First, it demonstrates a working prototype that transforms the earlier theoretical proposal into a usable educational web system. Second, it provides an initial pilot evaluation showing how student attempt logs can support adaptive diagnostics and personalized recommendation generation. Although the study is preliminary in scope, it demonstrates that adaptive web-based UNT preparation can function as a practical educational tool rather than only a conceptual model.

II. System Overview

The proposed system is a prototype adaptive web platform designed to support preparation for Kazakhstan’s Unified National Test through structured testing, performance tracking, and personalized feedback [3]. Unlike conventional exam-preparation tools that provide identical materials to all learners, the platform collects student interaction data and transforms it into topic-level diagnostics for adaptive support.

At the current stage, the prototype focuses on mathematics as the initial subject domain. This choice is justified by the structured nature of the UNT mathematics exam, which includes multiple interconnected topics and different difficulty levels. Such an organization makes mathematics suitable for topic-level analysis rather than simple total-score reporting.

The system consists of three main layers: the client layer, the backend layer, and the data layer. The client layer is implemented as a responsive web application that supports topic-based quizzes, progress tracking, and feedback delivery. The backend layer processes requests, serves assessment content, stores student responses, and prepares the data for further analytics. The data layer stores question metadata, attempt histories, and aggregated performance indicators. Together, these components provide the foundation for adaptive recommendation logic.

The user workflow is straightforward. A student selects a test or practice activity, answers mathematics questions, and receives a structured summary after completing the attempt. During the process, the system records topic information, correctness indicators, and timing-related data. These records are then aggregated into topic-level performance summaries that make it possible to identify strengths and weaknesses more precisely than by using an overall score alone.

A key feature of the prototype is its emphasis on structured logging. Instead of treating each attempt as an isolated test session, the platform stores measurable indicators such as topic accuracy, number of answered questions, difficulty-specific results, and average time per question. These indicators form the basis of the adaptive logic described later in the paper. Thus, the prototype serves not only as a testing interface but also as a practical environment for collecting and analyzing student performance data for personalized UNT preparation.

III. Pilot Study Design

To conduct the preliminary evaluation of the prototype, a pilot dataset of student attempt logs was used. Participation was anonymous and voluntary, and all participants agreed to the collection of assessment data for research purposes. The dataset contained 1,978 topic-level records representing 225 assessment sessions across 15 student profiles, with 15 attempts per student profile. Each assessment session was decomposed into topic-based performance entries, which allowed the system to analyze not only overall results but also fine-grained variation across mathematics topics. In total, the pilot covered 9 mathematics topics: numbers, equations, inequalities, systems, sequences, analysis, plane geometry, solid geometry, and vectors. This structure was consistent with the mathematics-oriented scope of the prototype and with the broader dissertation focus on topic-level personalization in UNT preparation.

The use of topic-level records was an intentional design choice. In traditional exam preparation systems, performance is often reduced to a single total score per test. However, such aggregation provides limited diagnostic value because students may perform well overall while still having serious weaknesses in specific areas. In contrast, the pilot design in this study treated each attempt as a structured set of topic observations. This made it possible to examine how a learner performed across multiple mathematical domains within and across attempts, which is more appropriate for adaptive support and recommendation generation.

Each record in the pilot dataset included several variables describing learner performance within a particular topic during a particular attempt. These variables included the number of answered questions in the topic, the number of easy, medium, and hard questions encountered, the number of correct answers by difficulty level, the total number of correct answers in the topic, the overall topic accuracy, the topic accuracy for each difficulty level, and the average time per question. As a result, the dataset supported both performance-based analysis and simple behavioral analysis through timing information. This is aligned with the dissertation architecture, where the platform is expected to log responses, correctness, timestamps, and aggregated mastery-related indicators for future adaptive analytics.

From the perspective of scale, the pilot dataset contained 7,587 answered questions in total across all topic-level records. On average, each topic-level record represented approximately 3.84 answered questions, while the mean topic accuracy across the dataset was approximately 0.69. The average response time per question was about 5.27 time units as recorded in the dataset. Although this pilot does not yet represent large-scale deployment, it is sufficient for evaluating whether the platform can ingest structured attempt data, compute topic-level indicators, and produce meaningful adaptive diagnostics.

The pilot study was not intended as a full experimental comparison between treatment and control groups. Instead, its purpose was feasibility-oriented. More specifically, the study aimed to answer the following practical questions:

1. Can the prototype consistently record and organize student attempt data at the topic level?
2. Can the logged indicators be transformed into interpretable performance summaries?
3. Can the resulting summaries be used to identify weaker and stronger areas for each learner profile?
4. Can these diagnostics support the generation of adaptive recommendations for further preparation?

This feasibility-oriented design is important because the present work represents a transition from conceptual architecture to applied implementation. In the earlier conference publication, the adaptive UNT platform was discussed mainly as a theoretical and survey-supported concept. By contrast, the current pilot focuses on whether the prototype can operate as a real analytical environment that connects testing, structured logging, and personalized diagnostics.

To preserve interpretability, the pilot evaluation in this paper does not rely on complex black-box models. Instead, it uses structured indicators derived directly from the attempt logs. This is consistent with the current stage of the dissertation project, where the platform is first being validated as a usable prototype and where mathematics serves as the initial subject for implementing adaptive logic. The emphasis, therefore, is not yet on maximizing predictive complexity, but on establishing that the prototype can support practical educational diagnostics in a stable and understandable way.

Overall, the pilot study design provides a controlled and analyzable setting for evaluating the feasibility of adaptive UNT preparation through a web-based platform. By organizing assessment data into topic-level records and using interpretable indicators of performance and timing, the study creates a practical basis for the adaptive logic described in the next section.

IV. Adaptive Logic

The adaptive logic of the prototype transforms student attempt logs into interpretable recommendations for further UNT preparation. At the current stage, the system uses a transparent topic-level analytical approach rather than a fully autonomous black-box recommendation engine. This choice is important because educational recommendations should remain understandable for both students and instructors [5].

After each attempt, the platform aggregates indicators separately for each mathematics topic. As a result, the learner is described not by a single overall score, but by a topic-specific performance profile. This is more useful for educational support, since a student may perform adequately overall while still showing clear weaknesses in particular domains such as vectors, inequalities, or systems of equations.

For each topic, the system computes a set of indicators that summarize achievement and response behavior. These include overall topic accuracy, number of answered questions, accuracy on easy questions, accuracy on medium questions, accuracy on hard questions, and average time per question. Together, these variables provide the basis for diagnosing topic mastery.

Based on these indicators, the platform assigns one of three diagnostic labels: weak, normal, or strong. A weak label reflects low overall accuracy, especially when it is combined with poor medium- and hard-level performance and relatively slow response time. A normal label indicates partial or developing mastery. A strong label reflects stable and comparatively confident performance across the topic. These labels are not permanent learner characteristics; they are current diagnostic states based on recent attempt data.

The recommendation layer then uses these labels to generate targeted feedback. If a topic is classified as weak, the system recommends revisiting core concepts and completing additional practice at foundational and medium levels. If a topic is classified as normal, the recommendation focuses on reinforcement and deeper practice. If a topic is classified as strong, the learner can be directed toward harder tasks, mixed-topic practice, or maintenance review. In this way, the prototype moves beyond simple result reporting and provides actionable study guidance.

A key advantage of this adaptive logic is explainability. The system can justify its recommendations using observable indicators such as low topic accuracy or weak hard-level performance. This makes the adaptive component suitable for prototype-stage educational use and creates a practical basis for later integration of more advanced machine learning models.

V. Results

The pilot dataset shows that the prototype can support structured topic-level diagnostics rather than only overall score reporting. The analysis covered 1,978 topic-level records derived from 225 assessment sessions completed by 15 student profiles.

The results revealed non-uniform performance across mathematics topics. The mean topic accuracy across the dataset was approximately 0.69, but the values differed by domain. The strongest topic was sequences, with a mean accuracy of about 0.73. It was followed by inequalities at about 0.72 and systems at about 0.71. In contrast, the lowest average performance was observed in numbers, solid geometry, and vectors, where the mean topic accuracy remained close to 0.67. These differences suggest that student achievement was uneven across the mathematics curriculum and confirm the value of topic-aware diagnostics.

Difficulty-specific analysis showed a clear pattern. Accuracy on easy questions was the highest in almost all topics, typically ranging from about 0.76 to 0.83. Medium-difficulty accuracy was lower, usually between about 0.62 and 0.70. Hard-question accuracy was the weakest dimension in nearly all cases and varied more substantially by topic. In particular, hard-level performance was relatively weak in numbers and solid geometry, where the mean hard-question accuracy was below 0.43. This pattern indicates that total topic accuracy alone is not sufficient for adaptive decision-making, since difficulty-specific results provide additional information about the depth of student understanding.

The longitudinal analysis also showed an upward trend across repeated attempts. When the average topic accuracy was aggregated by attempt number, the mean value increased from approximately 0.62 in the first attempt to approximately 0.76 in the fifteenth attempt. At the individual level, 13 out of 15 student profiles showed a positive difference between their first and final recorded attempt, while only 2 profiles showed a slight decrease. The average improvement between the first and last attempt was about 0.136 accuracy points. Although the pilot was not designed as a formal causal experiment, these observations suggest that repeated use of the platform may support progressive performance development over time.

The behavioral dimension of the data also contributed to the analysis. The average response time per question across the pilot was approximately 5.27 time units. In weaker topics, lower accuracy often co-occurred with relatively slower answering, which suggests hesitation or incomplete confidence. This supports the inclusion of timing data in the adaptive logic, since low performance combined with slower responses is more informative than correctness alone.

Overall, the results indicate that the prototype can distinguish between stronger and weaker student areas in a practically meaningful way. It can identify topics that require revision, detect differences between difficulty levels, and provide a data-based foundation for targeted recommendations. These findings support the feasibility of using the platform as a practical tool for personalized UNT mathematics preparation.

VI. Discussion and Conclusion

The pilot study shows that the proposed adaptive web platform can function as a practical tool for personalized UNT preparation at the prototype level. Its main contribution lies not in presenting a fully mature intelligent tutoring system, but in demonstrating that the transition from conceptual design to operational implementation is feasible. Unlike the earlier theoretical stage, the current study shows that the platform can record student attempts, aggregate topic-level indicators, and generate interpretable adaptive diagnostics from real usage data.

One of the main findings is that UNT preparation benefits from topic-level analysis rather than overall score reporting alone. The results showed that student performance was uneven across

mathematics topics and difficulty levels, which confirms the value of adaptive diagnostics. A learner may achieve an acceptable total score while still showing clear weaknesses in areas such as vectors, solid geometry, or numbers. Therefore, topic-based aggregation provides a more informative basis for guidance and recommendation generation.
Another important result is the usefulness of transparent adaptive logic. The prototype relies on interpretable indicators such as topic accuracy, difficulty-specific accuracy, question count, and average time per question. This makes the recommendation process understandable for students and instructors, which is especially important in educational settings where trust in automated suggestions may be limited. Even without a highly complex AI model, the system is already capable of supporting early diagnosis of learning gaps and more targeted preparation.
At the same time, the study has several limitations. The pilot dataset is relatively small, the evaluation is limited to mathematics, and no control group was used. For this reason, the observed improvement across repeated attempts should be interpreted as a promising pilot tendency rather than definitive proof of causal educational effectiveness. In addition, the current adaptive layer remains lightweight and should be expanded in future work through larger datasets, additional UNT subjects, and more advanced machine learning models.
Overall, the study confirms the feasibility of integrating adaptive analytics into UNT preparation through a web platform. The developed prototype demonstrates that student interaction data can be transformed into meaningful topic-level diagnostics and personalized guidance. Thus, even at an early stage, the platform provides a practical foundation for future large-scale development of machine-learning-enhanced exam preparation systems in Kazakhstan.

References List:

1. Bakas uulu, B., Smagulov, Y. *Analysis of dynamics of high school graduates who participated in the unified national test Kazakhstan*. International Electronic Journal of Mathematics Education, 2016. Vol. 11, No. 8. - Pp. 3176-3186.

2. Zamirbekkyzy, M., Bulakbay, Z. M. *Unified national testing: pathway to higher education*. Bulletin of Kazakh National Women’s Teacher Training University, 2023. No. 2. - Pp. 31-43.

3. Kem, D. *Personalised and adaptive learning: Emerging learning platforms in the era of digital and smart learning*. International Journal of Social Science and Human Research, 2022. Vol. 5, No. 2. - Pp. 385-391.

4. Ennouamani, S., Mahani, Z. *An overview of adaptive e-learning systems* - 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 2017. - Pp. 342-347.

5. Essa, S. G., Celik, T., Human-Hendricks, N. E. *Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review* . IEEE Access, 2023. Vol. 11. - Pp. 48392-48409.

6. *Спецификация теста по предмету «Математика» для Единого национального тестирования*. Министерство образования Республики Казахстан. https://testcenter.kz/wp-content/uploads/2025/10/10_%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0_%D1%80%D1%83%D1%81.pdf

7. Ospan, A., Myrzakerimova, A. *Designing an Adaptive Educational Platform for UNT Preparation: A Machine Learning-Based Approach*. 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST). Astana, 2025. 10.1109/SIST61657.2025.11139237