ПЕДАГОГИКА ҒЫЛЫМЫНЫҢ ТАРИХЫ ЖӘНЕ ЗАМАНАУИ БІЛІМ БЕРУ МЕН ҰЛТТЫҚ ТӘРБИЕНІҢ ӘДІСНАМАЛЫҚ МӘСЕЛЕЛЕРІ
ИСТОРИЯ И СОВРЕМЕННАЯ МЕТОДОЛОГИЯ ПЕДАГОГИЧЕСКОЙ НАУКИ, ОБРАЗОВАНИЯ И НАЦИОНАЛЬНОГО ВОСПИТАНИЯ

IRSTI 14.35.09 https://doi.org/10.51889/2959-5762.2025.88.4.001

Baizhanov N.,[1] Hao J.,[2] Yelemessov S.,[3] Tursynbekova A.[3]*

[1] National Testing Centre, Ministry of Science and Higher Education, Astana, Kazakhstan
[2] Educational Testing Service, Princeton, New Jersey, USA
[3] National Testing Center, Astana, Kazakhstan

REFORMING KAZAKHSTAN'S UNIFIED NATIONAL TEST FOR THE 21ST CENTURY: ACHIEVEMENTS, CHALLENGES, AND A NEXT-GENERATION ASSESSMENT MODEL

Abstract

Since 2004, Kazakhstan's Unified National Test (UNT) has played a crucial role in ensuring fair and transparent college admissions. The country's new goal of becoming a regional center for higher education and academic excellence, on the other hand, makes it clear that the current UNT needs to be reevaluated and revised. This paper examines what UNT has done well and what it has not done sufficiently to meet the new needs. It gives an idea of what a next-generation UNT could look like, based on best practices from around the world. Given recent global and national evidence on twenty-first-century skills, we consider one possible pathway for gradual enhancement of the UNT that preserves its equity role by testing modular, skills-focused components and using human-in-the-loop assessment. This proposal is exploratory and aims to foster dialogue among stakeholders.The suggested solution employs a modular and competency-based framework that incorporates Skill-Based Tests (SBT) to assess critical thinking, academic writing, research, communication, AI, digital literacy, and creative problem-solving.

Keywords: Unified National Test, National educational assessment reform, 21st-century skills, Artificial intelligence in assessment, human capital development, higher education policy, test modernization.

Н.А.Байжанов,[1] Хао Цзянган,[2] С.М.Елемесов,[3] А.Е.Турсынбекова [3]*
[1] Ұлттық тестілеу орталығы, Ғылым және жоғары білім министрлігі, Астана қ., Қазақстан
[2] Educational Testing Service, Принстон, Нью-Джерси қ., АҚШ
[3] Ұлттық тестілеу орталығы, Астана қ., Қазақстан

ҚАЗАҚСТАННЫҢ ҰЛТТЫҚ БІРЫҢҒАЙ ТЕСТІЛЕУІН XXI ҒАСЫРҒА БЕЙІМДЕУ: ЖЕТІСТІКТЕР, МӘСЕЛЕЛЕР ЖӘНЕ ЖАҢА БУЫН БАҒАЛАУ МОДЕЛІ

Аңдатпа

2004 жылдан бастап Қазақстандағы Ұлттық бірыңғай тестілеу (ҰБТ) жоғары оқу орындарына қабылдауда ашықтық пен әділдікті қамтамасыз етуде маңызды рөл атқарып келеді. Алайда елдің жоғары білім мен академиялық сапа саласында өңірлік хабқа айналу жөніндегі жаңа мақсаты қолданыстағы ҰБТ моделін қайта қарауды талап етеді. Бұл мақалада ҰБТ жетістіктері мен оның қазіргі талаптарға сәйкессіздіктері талданады. Авторлар халықаралық үздік тәжірибелерге сүйене отырып, болашаққа бағытталған жаңа буын ҰБТ моделін ұсынады. XXI ғасыр дағдылары жөніндегі жаһандық және ұлттық деректерге сүйене отырып, біз ҰБТ-ны біртіндеп жетілдірудің, оның теңдік қызметін сақтай отырып, дағдыға бағытталған модульдік компоненттерді сынақтан өткізу және адам қатысатын бағалауды (human-in-the-loop) қолдану арқылы іске асатын ықтимал жолдарының бірін қарастырамыз. Бұл ұсыныс зерттеушілік сипатта және мүдделі тараптар арасында пікірталас өрбітуге бағытталған. Бұл модель модульдік құрылымға және дағдыларға негізделген тесттерге (Skill-Based Tests – SBT) сүйенеді, олар сыни ойлау, академиялық жазу, зерттеу, коммуникация, цифрлық және жасанды интеллект сауаттылығы мен креативті тапсырмаларды шешу дағдыларын бағалауға бағытталған.

Түйін сөздер: Ұлттық бірыңғай тестілеу, ұлттық бағалау жүйесін реформалау, XXI ғасыр дағдылары, бағалаудағы жасанды интеллект, адами капиталды дамыту, жоғары білім саясаты, тесттерді жаңғырту.

Байжанов Н.А.,[1] Хао Цзянган,[2] Елемесов С.М.,[3] Турсынбекова А.Е.[3]*
[1] Национальный центр тестирования, Министерство науки и высшего образования, г.Астана, Казахстан
[2] Educational Testing Service, Принстон, г.Нью-Джерси, США
[3] Магистр, Национальный центр тестирования, г.Астана, Казахстан

РЕФОРМИРОВАНИЕ ЕДИНОГО НАЦИОНАЛЬНОГО ТЕСТА КАЗАХСТАНА В КОНТЕКСТЕ XXI ВЕКА: ДОСТИЖЕНИЯ, ВЫЗОВЫ И МОДЕЛЬ ОЦЕНКИ НОВОГО ПОКОЛЕНИЯ

Аннотация

С 2004 года Единое национальное тестирование (ЕНТ) в Казахстане играет ключевую роль в обеспечении справедливости и прозрачности при поступлении в вузы. Однако новая стратегическая цель страны, а именно стать региональным центром высшего образования и академического превосходства, что требует пересмотра и модернизации существующей модели ЕНТ. В данной статье рассмотрены достижения системы и выявлены её ограничения в условиях новых вызовов. Авторы предлагают модель ЕНТ нового поколения, основанную на международных подходах и компетентностной структуре. Учитывая недавние международные и национальные данные о навыках XXI века, мы рассматриваем один из возможных путей постепенного совершенствования ЕНТ, сохраняющий его функцию обеспечения равенства, посредством апробации модульных, ориентированных на навыки компонентов и использования оценивания с участием эксперта (human-in-the-loop). Это предложение носит исследовательский характер и направлено на стимулирование диалога между заинтересованными сторонами. Особое внимание уделяется внедрению модульных компонентов и тестов на основе навыков (SBT) для оценки критического мышления, академического письма, исследовательских и коммуникативных умений, цифровой и ИИ-грамотности, а также творческого решения проблем.

Ключевые слова: Единое национальное тестирование, реформа национальной системы оценивания, навыки XXI века, искусственный интеллект в оценке, развитие человеческого капитала, политика в сфере высшего образования, модернизация тестов.

Introduction. The Unified National Test (UNT) in Kazakhstan has been a key component of the country's university admissions system since its establishment in 2004. It eliminated entrance exams administered by different schools and replaced them with a single standardized test. The UNT made access to higher education more fair, open, and trusted by the public by closing the gaps between urban and rural areas and reducing bias in the selection process [1]. The UNT also served as a guiding framework that effectively forced secondary schools to align their curricula with clearly defined performance targets. This alignment altered how teachers taught and helped students learn more effectively [2].

Over the last twenty years, the UNT has gradually improved. This includes transitioning to computer-based delivery, covering more subjects, and introducing more challenging item types that require in-depth analysis. National education policymakers and international partners, such as the World Bank, have helped ensure these changes maintain the exam's psychometric soundness and consistency in relevance to current education [3; 4].

The UNT has had significant policy successes since its introduction in 2004. It replaced fragmented university entrance exams with a single, standardized assessment, bringing fairness, transparency, and objectivity to the admissions process [1]. The system helped rural applicants compete fairly with urban applicants by lowering regional differences [2]. The use of strict test security measures, automated scoring, and centralized administration made it much less likely that people would accidentally interfere with or corrupt the system. This made people trust that the system was fair and based on merit. The UNT changed how teachers taught over time by providing them with a straightforward way to ensure that school curricula met national education standards. This had a significant washback effect, which helped make the quality of education more consistent nationwide. Over the years, minor improvements have strengthened the structure and delivery of the UNT. The National Testing Center (NTC) introduced computer-based testing and digital registration and began piloting open-ended questions. A review funded by the World Bank suggested additional technical improvements, including the addition

of more items and enhancements; however, its current design does not align with Kazakhstan's evolving needs. The UNT has made university admissions fairer. It aims to become a leader in higher education and innovation in the region, so the UNT needs to be updated to meet the new needs for identifying and hiring the proper personnelher education and innovation in the area, so the UNT needs to be updated to meet the new needs for finding and hiring the right people.

However, recent events have shown that the UNT is not in line with the changing needs of modern, rigorous higher education programs. For instance, in 2024, one-third of the applicants who received funding from UNT did not meet the academic and language requirements of the new Heriot-Watt University branch in Aktobe, despite having high UNT scores [5]. This accident triggered important questions about the UNT's effectiveness in accurately assessing university readiness, especially for English-medium programs that emphasize critical thinking.

Simultaneously, Kazakhstan's vision to become a regional center for an innovation-driven economy and a hub of higher education calls for a more comprehensive shift in UNT's assessment focus. Measuring 21st-century competencies, such as problem-solving, communication, digital literacy, and creativity, has become essential for global competition and the digital economy [6; 7]. The UNT should be reformed to align talent attraction and selection with national priorities so that, beyond its admissions function, it can operate "as a strategic mechanism for developing human capital" [8, p. 7], aligned with a knowledge-based economy and implemented with strong safeguards for fairness and transparency [8, p. 5].

International studies emphasize that national high-stakes exams must evolve toward assessing higher-order and transversal skills to remain relevant in a rapidly changing global economy. The World Bank's 2023 strategic review of the UNT highlights that although the test has strong foundations of fairness and psychometric stability, it "does not yet sufficiently capture critical thinking, academic writing, or digital competencies demanded by modern higher education" [3, pp. 42–45]. Similar concerns are reflected in OECD reports, which stress that 21st-century systems should evaluate "problem-solving, creativity and digital literacy as core components of student readiness" [7; 10]. Research on Kazakhstan's higher education also notes gaps between school-based preparation and university expectations, particularly in academic writing and argumentation, which remain "weakly developed among first-year students" [9, p.101]. Global frameworks, such as the PISA Creative Thinking assessment, argue that evaluating innovative and analytical skills provides more accurate signals of learner potential compared to rote-based models [10]. Taken together, these findings suggest that while the UNT has ensured equity and transparency, its design must align with emerging international standards to better assess the competencies required for success in contemporary university programs and knowledge-based economies.

Basic provisions. Limitations of the UNT. The UNT has made university admissions fairer. As Kazakhstan aims to become a leader in higher education and innovation in the region, the UNT needs to be updated to meet new requirements for identifying and selecting the right personnel. Here are the main points of the limitations:

First, the UNT has been crucial in ensuring that everyone has equal access to higher education; however, its current design does not yet fully reflect the range of skills necessary to succeed in the 21st century. The current system does not adequately assess important skills such as critical thinking, creativity, communication, and digital literacy, which are widely regarded as essential for success in both school and the workplace. Many first-year college applicants struggle with tasks such as writing arguments and logical reasoning [9]. This suggests that there may be a gap between what is tested and what is required for college-level study. Filling this gap presents an opportunity to align assessments more closely with national education goals and international competency frameworks.

Second, the way the UNT is set up still puts most of the emphasis on memorizing facts and not enough on higher-order thinking skills. The test meets the standards for the curriculum, but it does not yet have good ways to find out how well applicants can organize information, come up with new ideas, or make structured arguments. In a knowledge economy and higher education that is changing quickly,

these cognitive skills are becoming more and more important for success. People could learn transferable skills like creative thinking, analytical reasoning, academic writing, and others that are important for lifelong learning and being able to adapt to new situations at work, if the assessment model were even better.

Third, the design and implementation of the UNT do not yet fully use modern psychometric methods like item response theory (IRT), large-scale pretesting, and statistical equating. Most people agree that these are the best ways to make sure that assessment results are reliable, valid, and fair. The current test format also does not provide applicants, teachers, and schools with sufficient diagnostic feedback, making it less effective as a formative tool for improving instruction. Improving the psychometric sophistication of the UNT would bring Kazakhstan's testing system up to international standards, making people more confident in its results.

Lastly, even though recent changes have made the UNT more flexible (for example, by allowing students to choose their subjects and take the test multiple times a year), the overall structure remains essentially unchanged. It does not fully illustrate the differences between various university programs and their students. The current model does not yet allow for modular alignment with interdisciplinary or competency-based tracks, which are becoming increasingly common in Kazakhstani universities. As a result, applicants may be evaluated in ways that do not fully align with the needs of the academic paths they wish to pursue. Colleges and universities can also tailor their course requirements to fit their course structures better. By addressing these issues, the UNT ensures that college admissions are fair and relevant, and it supports innovative approaches that do not alter their entrance requirements to fit their course structures better, which remains an integral part of Kazakhstan's college admissions system. However, recent events, such as the well-publicized problems with admissions at the Heriot-Watt University branch in Aktobe, demonstrate the importance of ongoing changes. A redesign of the UNT that looks to the future and takes into account global best practices and national priorities would make it better equipped to serve both educational equity and excellence in a world where the economy and education are rapidly changing.

Proposed Reform Directions. The World Bank (2023) recommends the use of open-ended questions, new psychometric methods, diverse content types, and enhanced feedback systems [3]. The OECD (2023) also states that national assessment systems should evolve into next-generation assessments that reflect real-world skills, utilize digital tools, and ensure they are valid, comparable, and of high quality, aligning with international frameworks [10; 11].

The reform proposal is based on these ideas and has these goals:

- Focus on the most important strengths, such as being fair, ensuring everyone is treated equally, and selecting students based on their skills.
- Enhance the content with advanced skills, such as creativity, digital literacy, and critical thinking.
- Use item response theory, statistical equating, and large-scale item pretesting to make psychometric practices more up-to-date and make test results more valid and reliable.
- Make a modular structure that has both required and optional parts that are based on the applicants' career goals and academic interests.
- Use a modular structure that includes both required and optional components, tailored to the academic and career goals of the applicants.
- Use AI to help with building tests, scoring them automatically, and making detailed profiles of applicants to help make better and more complete decisions about who to admit to college.
- Implement changes in stages by testing new ideas, training all involved parties, and ensuring that all applicants have equal access.

These pieces work together to form the foundation of a new type of UNT model. The main goal is to change the UNT from a regular selection tool into a strategic tool that helps Kazakhstan’s national strategy for educational excellence, innovation, and global competitiveness.

Materials and Methods. To come up with a plan to change Kazakhstan’s UNT, this study uses a qualitative, mixed-methods approach that includes policy analysis, a literature review, and comparative

education research. We read national policy papers like the Government's Concept for Higher Education and Science Development (2023–2029) and the Ministry of Education's rules to find out more about Kazakhstan's education goals and rules [12].

Next, we examined the World Bank's (2023) strategic review of the UNT and identified significant issues with the system, along with recommendations for addressing them [3]. We compared these ideas to the best practices for assessment worldwide, focusing on how to measure 21st-century skills [10; 11], digital testing trends, and adaptive formats. The SAT/ACT in the U.S., the A-levels in Britain, the gaokao in China, and holistic admissions models were all used as examples. We also reviewed frameworks such as PISA and the GRE, which assess competencies like creative thinking and analytical writing skills, which have been shown to correlate with college success [13].

To ensure contextual relevance, we incorporated secondary stakeholder feedback, including faculty concerns about the readiness of UNT graduates (e.g., the Heriot-Watt Aktobe case) and lessons from pilot studies on digital testing and AI-based scoring, as reported by the World Bank and the NTC. The reform model was developed through an iterative design process, where initial ideas were refined through conceptual analysis and benchmarking against global trends. Although primary data collection was beyond the scope of the study, proposals are informed by credible secondary sources and previous fieldwork.

Results. New Approach and New Modules of the Next-generation UNT. The primary objective of this study is to develop a design for the next generation of the UNT that is modular, competency-based, and aligned with current international testing standards. The main idea behind this design is a modular exam structure that is different from the current UNT's one-size-fits-all approach. In the reformed model, every applicant would undergo subject-specific and skill-specific modules, which would be mixed according to their intended field of study requirements. While preserving comparability and fairness across multiple test versions, this flexible structure allows the assessment to be customized and relevant to applicants' academic goals.

Under the proposed new UNT model, all applicants would complete a set of core modules that remain obligatory. These include History of Kazakhstan (a civic knowledge test, potentially scored on a pass/fail basis for state grant eligibility) and one or more specialized subject modules corresponding to the applicant's chosen discipline. For example, an applicant aiming for an engineering career might take advanced Mathematics and Physics modules, while one headed for a medical career would take Biology and Chemistry modules. The content of these subject modules would be drawn from the high school curriculum but updated to include more application-based and analytical questions (discussed further below). In addition to the core subject exams, applicants would undertake a series of Skill-Based Test (SBT) modules that assess transversal skills important for university success and beyond. The initial suite of SBT modules proposed includes Academic Writing and Research Skills (AWRS), measuring the ability to produce organized, evidence-supported writing and understand scholarly texts; Critical Thinking and Logic (CTL), assessing reasoning, argument evaluation, and problem-solving abilities; Communication Skills (CS), evaluating listening comprehension and oral communication, including an element of spoken interaction; Artificial Intelligence (AI) and Digital Literacy (ADL), gauging digital competencies, computational thinking, and basic knowledge of AI and data use; and Creative and Design Thinking (CDT), examining creativity, innovation in problem-solving, and design-oriented thinking. The relative emphasis and proportional weighting of these modules in the next-generation UNT is presented in Figure 1.

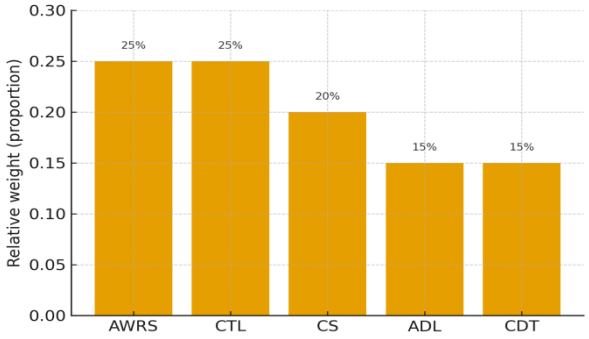

Figure 1. Distribution of relative weights across SBT modules

The first three modules, AWRS, CTL, and CS, are conceptually grouped as an English for Academic Purposes (EAP) module when delivered in English, reflecting an integrated approach to academic language and skills. Indeed, the EAP module is envisioned as an early focus: it would be initially implemented in English to serve the needs of international branch universities and competitive programs and later adapted into Kazakh and Russian. The ADL and CDT modules cover modern skill areas that are not covered by traditional subjects. By the time Phase III of the reform is over (full implementation), all applicants should have finished a mix of subject and skill modules. The Ministry of Education would probably require specific modules (like core subjects plus one or two key skills) for students who want to get into a state-funded school, making sure that they are in line with national goals. At the same time, universities could have leeway to require additional modules or set their thresholds for specific programs (e.g., an art and design program might place greater weight on CDT scores, while an IT program might place greater weight on ADL and math). Crucially, every configuration of modules would carry the same total weight in admissions decisions to remain fair; the differences lie in content emphasis rather than total exam load.

This modular approach offers several benefits: it introduces flexibility and differentiation into the UNT while preserving its role as a standard national exam. The overall configuration of this architecture is illustrated in Figure 2, which presents how core subject modules and SBT modules are combined into customized exam paths across different academic directions. Applicants can highlight areas relevant to their future studies where they excel, and colleges gain more comprehensive information on candidates' competencies. Crucially, every module, including a critical thinking exercise or a mathematics test, will be built to be psychometrically equivalent in rigor. Strict calibration and equating procedures (discussed later) will ensure that, regardless of which combination of modules a student takes, the scores are comparable and can be aggregated or weighed fairly in making admissions and scholarship decisions. In practice, an admissions committee typically reviews a profile for each applicant, which typically includes subject scores, skill module scores, and a composite score, depending on the application's requirements. This richer profile moves beyond the single composite UNT score currently used. It allows, for instance, an applicant with average subject scores but an outstanding writing score to still be considered for a humanities program or an applicant with top math or science scores but lower writing skills to be admitted to an engineering program that can support the development of those writing skills.

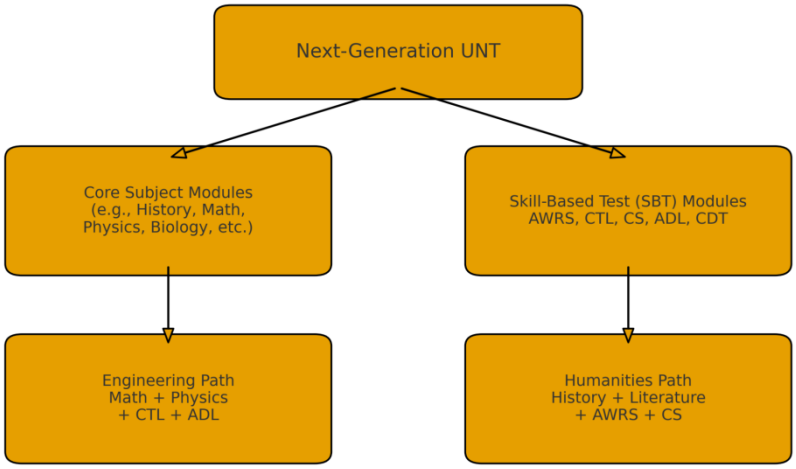

Figure 2. Modular structure of the next-generation UNT

By embedding SBT modules into the exam, Kazakhstan squarely addresses the skills gap identified in its applicants. Each new module has a clear justification. The AWRS component responds to observations that many entrants lack structured academic writing ability – a skill critical for term papers and research projects in university [9]. Including a writing task (e.g., an argumentative essay or a research interpretation exercise) in the UNT will signal to schools the importance of emphasizing writing instruction and, over time, should reduce universities' need to provide remedial writing courses. The CTL module targets inference, logical reasoning, and analytical skills, drawing on global models such as the Watson-Glaser Critical Thinking Appraisal [14] and research linking critical thinking to college success [15]. The UNT will incentivize teaching methods that develop reasoning by having test-takers identify assumptions, evaluate arguments, or solve novel problems. The CS module stresses how important it is to be able to communicate well, which is a skill that some schools do not teach well enough. It probably means listening to part of an academic lecture and then giving a structured oral response or taking part in a recorded interview. To make this module more like how people talk to each other these days, we suggest adding digital communication elements, like using AI-powered interactive prompts or interpreting information from digital sources. Together, AWRS, CTL, and CS (the EAP suite) reflect competencies tested in some form by many leading universities via essays or interviews; their inclusion aligns Kazakhstan's admissions with those international practices [11; 10].

An original addition that directly addresses Kazakhstan’s strategic focus on digital transformation and prepares applicants for a technologically rich environment is the AI and ADL module. This module would evaluate knowledge of ethical concerns in AI , internet research savvy, basic coding logic, and algorithmic thinking. By introducing ADL, Kazakhstan pioneers the assessment of digital literacy in a high-stakes exam, acknowledging that familiarity with AI and digital tools is now as fundamental as literacy and numeracy for the 21st century (16; 17). The CDT module is similarly forward-looking. Measuring creativity in standardized tests is a challenging task. Nonetheless, the module can include open-ended tasks where applicants must devise a solution or design that meets specific requirements or respond to a hypothetical situation in a novel way. Recent international efforts, such as the PISA 2022 Creative Thinking test, provide us with ways to evaluate these ideas [10] fairly. The CDT module's presence at UNT will demonstrate to teachers and students that creative problem-solving and innovative ideas are valued. Schools can incorporate more project-based and exploratory learning if they devote even a little focus to creativity. Even a slight emphasis on creativity can lead to more project-based and exploratory learning in schools. Zhao (2018; 2019) says that putting too much emphasis on tests can stifle creativity. In contrast, Kazakhstan's decision to incorporate creativity into its admissions exam could help reverse this trend and foster a more creative academic culture [18; 19].

Initially, the SBT modules could be added to the main UNT as optional or pilot components to collect data and familiarize people with them. The sequential rollout of these changes - Phase I pilot, Phase II scaling, and Phase III full integration - is summarized in Figure 3.

Figure 3. Phased implementation of the next-generation UNT reform

For example, during Phase I (pilot stage), applicants could volunteer to take an EAP module after their primary test and receive a certificate or bonus consideration for scholarships without it being mandatory. This approach allows stakeholders to become familiar with the new formats in a low-stakes way. Over subsequent phases, as confidence and proof of validity grew, the SBT modules would be scaled up, translated into Kazakh and Russian, and eventually integrated as compulsory elements of the UNT (Phase III full rollout). By Phase III, we anticipate that some modules might be combined or adjusted – e.g., the EAP components could remain bundled for efficiency – but the principle is that all applicants will be assessed not just on what they know but on how well they can write, think, communicate, and solve problems creatively, in addition to demonstrating digital literacy. The modular format also naturally supports a micro-credentialing approach; down the line, high performance in a module like ADL or CTL could be recognized with a separate certificate that applicants might use for other purposes (employment or later studies) beyond university entrance. Global trends toward credentialing discrete skills and creating student portfolios of competencies [11] speak to this.

All in all, the structure of the next-generation UNT is designed to be balanced and comprehensive. It dramatically expands the assessment by adding new ability dimensions while retaining subject tests to ensure comprehensive coverage of key academic knowledge. The result is a more complete profile of an applicant's readiness for higher education in the 21st century. Table 1 (hypothetical) would illustrate sample module combinations for different intended majors, but in prose: an engineering applicant might submit Math, Physics, AWRS, CTL, CS, or ADL; a journalism applicant might take History, Literature, AWRS, CS, CTL, perhaps an optional Creative module; an IT applicant: Math, Physics, CTL, ADL, etc.

Under this approach, universities can be assured that admitted applicants have met the required knowledge and skills, while applicants have several opportunities to demonstrate their strengths. Thus, the modular UNT is positioned as a flexible but standardized framework that combines the advantages of uniform assessment with the necessity of allowing different futures.

Modernizing Test Design and Content Quality. Along with the new modules, the proposed reform of Kazakhstan's UNT stresses a complete modernization of test design, content, and scoring system. The primary aim is to develop a fair, accurate, and valid assessment system that reflects the international standards and skills of the twenty-first century. Key elements include expanding item formats, adopting psychometric advancements, enhancing cognitive demand, and offering actionable feedback to applicants and institutions.

Combining performance-based and constructed-response items into skill-based (SBT) traditional subject modules is a core transformation. The updated UNT is currently primarily composed of multiple-choice questions. In Academic Writing and Research Skills (AWRS), for example, students will complete essays, while in STEM fields, they will engage in open-ended problem-solving exercises. In the social sciences, scenario-based assignments will be used. These tools are designed to encourage more critical thinking by allowing applicants to demonstrate their ability to communicate, reason, and solve problems effectively. According to the World Bank, this type of diversification aligns with practices in Turkey, China, and certain parts of Europe. Based on World Bank advice, this type of diversification aligns with global standards in countries such as Turkey, China, and parts of Europe.

Although it can be difficult to score open-ended items consistently, the benefits they provide in evaluating important skills outweigh the drawbacks. For a chemistry module, applicants might have to provide an explanation of experimental data. Because a history assignment might also call for looking at a historical source, the focus should shift from rote memory to applied knowledge.

Using modern psychometric methods is equally crucial, especially Item Response Theory (IRT), which forms the basis of cross-test equating and scoring. IRT makes it possible to calibrate items based on discrimination and difficulty parameters, which is different from traditional scoring models and leads to more accurate and equitable interpretations of applicant performance. Scores will be comparable across test versions and years thanks to anchor items and methodical equating. This ensures that applicants are evaluated against a consistent standard, eliminating inequalities caused by annual test variance. Furthermore, supporting ongoing quality control is regular pretesting of new products, either as unscored components or by pilot runs. Although adaptive testing is a long-term goal, the immediate focus is on using fixed, intelligently assembled forms.

Improving the cognitive rigidity of test materials is another pillar of the reform. UNT products are mostly arranged loosely according to difficulty, usually based on outward features. From knowledge recall to application, analysis, and evaluation, the proposed change toward Bloom's taxonomy allows a more sophisticated classification by cognitive process. Every subject module will thus have a range of assignments that require higher-order thinking. For instance, math tests might include real-world modeling problems, while biology might ask applicants to interpret experimental data. Emphasizing critical thinking and problem-solving, this type of alignment reflects trends in tests of world knowledge, such as PISA. Item writers will receive specialized training; test blueprints will require a percentage of higher-level tasks to ensure cognitive depth and greater candidate differentiation.

Equally vital is increasing content validity and fairness. Expanding the contributor base beyond a small group at NTC to include experienced teachers, university faculty, and international experts will diversify perspectives and enhance the quality of items. All items will be reviewed for curriculum alignment, clarity, and absence of bias, whether regional, cultural, or gender-related. Additionally, fairness across Kazakhstan's trilingual landscape (Kazakh, Russian, and English) will be addressed through parallel item development and cross-language equivalency checks. Localization efforts will make sure that all test versions are in the same language, but students will still be able to take the UNT in the language of instruction. For content to be accurate and comparable across computer-based testing (CBT) platforms, it is essential to invest in item banking systems and translation tools. Although the current system only accepts static item formats, Kazakhstan has already made CBT the accepted method for taking the UNT. It will be simpler to create tests that use interactive questions, simulations, and multimedia if CBT has more features. This will increase test taker engagement. People with disabilities who are applying for jobs benefit from it as well. To transition to more sophisticated international standards, Kazakhstan must further enhance its digital infrastructure. This entails constructing dependable high-speed networks and secure, technologically advanced testing facilities. Effectively. It also benefits applicants with disabilities. A full-scale, next-generation CBT system would allow for more advanced features, such as automated data capture of response timing and interaction patterns, even though preliminary steps have been taken through digital pilot initiatives. Test improvement, item diagnostics, and real-time analytics can all benefit from this data.

Additionally, CBT makes it possible to change the way that candidate evaluation and scoring are carried out. Although objective items can still be automatically scored using answer keys, more complex tasks like essays and oral presentations require a different approach. A hybrid scoring model balances efficiency and integrity by having AI perform the initial review and human raters look at complex or borderline cases. This two-tiered approach ensures scalable and dependable scoring.

Applicants will receive comprehensive reports detailing their performance in each module and skill area, as well as their overall scores. For example, a report might state that someone was proficient in algebra but struggled with geometry, or that an essay was well-structured but lacked sufficient evidence to support its assertions. In addition to providing feedback, these detailed profiles highlight each

applicant's strengths and areas for improvement. Schools can support student growth and make well-informed decisions with the use of this kind of information. Over time, gathered response data might also highlight curriculum misalignments or equity gaps, which would guide future policy changes.

In the end, Kazakhstan has a rare chance to revolutionize its national assessment ecosystem by updating the UNT's structure and content through improved CBT and AI integration. Adding new content modules alone will not bring Kazakhstan up to international standards. Other technical adjustments will also be required. The proposed changes, which include a wider range of item types, modern psychometrics, and AI-powered scoring and analytics, represent a systemic upgrade that enhances the system's validity, fairness, and transparency. These new features will make the UNT more useful and dependable, making it a key tool for promoting fairness and academic excellence in schools nationwide.

Integration of Artificial Intelligence in the UNT System. The strategic use of AI throughout all stages of the assessment lifecycle is a key part of Kazakhstan's next-generation UNT. Given the increasing complexity of the new UNT, which has a modular structure, diversified item types, and skill-based components, AI technologies are essential for maintaining efficiency, quality, fairness, and scalability. This section outlines AI's major areas: test assembly, item generation, scoring, test security, analytics, and system governance.

One of AI's earliest and most impactful applications will be in the intelligent assembly of test forms. The modular design of the new UNT requires numerous equivalent test forms adapted to candidates' intended fields of study (e.g., medicine, engineering, humanities). Traditionally, test form construction is a manual, time-consuming process. With AI-driven form generation, calibrated items can be algorithmically selected from a tagged item bank based on predefined blueprints. Every blueprint specifies cognitive level balance, skill coverage, and subject mix. A math module might require 30% algebra, 40% calculus, and 30% applied problems, with a designated distribution of difficulty levels. AI guarantees the constant meeting of all kinds of these limitations.

Customizing is another advantage of form assembly driven by AI. Before the exam, applicants may complete a pretest survey indicating their academic interests and study preferences. AI algorithms can then provide applicants with personalized tests, such as biology-heavy science questions for future doctors or logic-heavy tasks for future computer scientists. This change to the pre-assignment makes the test content more relevant while maintaining the comparability of the forms. Crucially, all variations will be psychometrically equated through Item Response Theory (IRT), using anchor items to link different forms onto a standard scale.

Equally transformative is the use of AI in item generation and quality control. Valid assessments depend on high-quality test items; however, creating them in large numbers is a considerable amount of work. Automated Item Generation (AIG) utilizes AI, specifically natural language processing and machine learning, to create draft items from templates or by modeling content. For example, an AI could modify the contextual variables of a critical thinking task to create multiple versions of it while maintaining the same cognitive demand. Expert item writers can then review, refine, and pilot these drafts.

Generative AI models such as GPT-4 can further assist by suggesting reading passages, brainstorming essay prompts, or generating distractor options for multiple-choice questions. These AI-generated materials require human review, but they significantly reduce development time and add variety to the item pool. Kazakhstan's NTC is already testing these kinds of tools, and early results show that AI can make assessment content more creative and increase its quantity. This method also makes tests safer because item banks that are updated often lower the chance of questions leaking or being seen too much.

Nevertheless, products created by AI still require examination. It is important to have a multi-stage validation process that includes expert checks, pilot testing, and bias screening. Items must be reviewed to ensure they align with the curriculum, are clear, suitable for individuals from diverse cultural and linguistic backgrounds, and present cognitive challenges. The system ensures its growth and validity using AI-driven innovation and human oversight.

AI also plays a critical role in scoring, particularly for constructed responses such as essays and oral answers. In the past, scoring such responses at scale relied entirely on human raters, expensive, slow, and prone to inconsistency. Today's AI scoring engines use machine learning models trained on human-annotated responses to evaluate content accuracy, coherence, language use, and structure. These engines have achieved performance levels comparable to those of human raters in international exams such as the GRE and TOEFL.

The UNT will adopt a hybrid scoring model: AI will handle initial scoring, with human raters auditing a sample of responses and reviewing those flagged as anomalous. This allows for a rapid turnaround, applicants can receive their scores within days, while preserving accuracy and fairness. Scoring rubrics will be standardized, and the system will be calibrated on Kazakhstani applicant data to ensure local relevance. When there is little trust or disagreement, people tend to make the final decision. Applicants can still appeal automated scores, and trained human experts will handle rescoring.

AI-driven scoring not only ensures efficient evaluation but also enables the generation of rich applicant profiles, supporting more informed and holistic admissions decisions. Applicants will get diagnostic reports that show their overall scores and sub-scores by module or skill area. For instance, candidates may discover that their essays were well-structured but lacked evidence, or that they did well in algebra but poorly in geometry. Although this kind of diagnostic report is rare in high-stakes exams, it can help with better preparation, learning, and even system-level curriculum modifications.

Maintaining test fairness will be more difficult as the UNT switches to computer-based testing (CBT). Through data forensics, biometric authentication, and automated proctoring, AI tools will contribute to test security. When computer vision algorithms are being tested, they can watch video feeds and identify odd events, such as people moving in odd ways, people who should not be there, or people who keep turning away. AI can also look for indications of potentially problematic behavior in the way applicants type, move their mouse, and respond to objects.

After the test, AI will help identify patterns that suggest cheating, such as candidates having answers that are too similar or response profiles that do not make sense (for example, getting hard questions right but missing easy ones). These forensic checks will help identify individuals who are working together, pretending to be someone else, or using unauthorized materials. Biometric tools, such as facial recognition at check-in, will prevent identity fraud. Global testing organizations already employ these security measures and will tailor them to Kazakhstan's national context. They are critical for building public trust in a digital, high-stakes assessment environment.

Another forward-looking application is AI-enabled data analytics for system monitoring and policy planning. The large volume of data generated by the new UNT, item-level responses, timing logs, and sub-scores offers a valuable resource for improving teaching, learning, and educational governance. AI algorithms can detect patterns across schools, regions, and demographic groups. For instance, unsupervised machine learning may reveal that high school applicants in some rural regions systematically underperform in digital literacy, prompting targeted interventions or teacher training.

AI can also help check how well UNT modules can predict the future. Models can find the sub-scores that best predict academic success by comparing UNT results to university performance data. These insights can help universities determine the appropriate weight to assign to various aspects of the application or how to manage admissions effectively. AI-powered dashboards can help the Ministry of Education monitor system-level trends over time, such as when critical thinking skills improve or when specific skill gaps become more pronounced. They can then change policies to reflect these changes.

All AI systems within the UNT ecosystem will be required to adhere to strict rules governing ethics and governance. A human-in-the-loop approach is important because AI can help and improve human decision-making, rather than replace it. Experts will regularly review AI outputs for content generation, scoring, and security analytics. Transparent communication with stakeholders, applicants, educators, and policymakers, will ensure that AI usage is understood, trusted, and contestable. Appeal mechanisms will be in place for any AI-related scoring or misconduct decisions.

Bias mitigation is a core priority. All AI systems will be evaluated for fairness across languages, genders, and geographic groups. For example, essay scorers will be audited to ensure they do not favor writing styles that are more common among urban applicants or native Russian speakers. Translation tools used in multilingual test forms will be tested for semantic equivalency. Regular quality assurance audits and calibration exercises will ensure that high standards are maintained across all functions.

AI integration is a cornerstone of the next-generation UNT. It supports scalability, customization, efficiency, and fairness, essential for a modern, skills-based national assessment system. AI can turn a regular test into a new, flexible, and policy-making educational tool when used with honesty, ethical safeguards, and expert supervision. Kazakhstan is leading global assessment reform because it is committed to this vision. The UNT is a model for fair, data-driven, and future-ready educational systems.

Discussion. The proposed changes to Kazakhstan's UNT are framed as a potential option for improvement, rather than an obligation. Importantly, any transformation of the UNT must remain socially sensitive: protect opportunities for rural and low-SES applicants; maintaining language parity; preventing financial strain on households; and being vigilant about any unintended effects on teaching (such as stifling creativity). In this setting, the UNT should "not be seen as a mere technical upgrade, but as a strategic necessity", and it can develop "into a strategic tool for developing human capital" (Baizhanov 2025), promoting national objectives without restricting access for diverse learners. Changes should be implemented gradually, following pilot testing and validation, with human oversight in assessment and systems in place to ensure fairness and the ability to appeal (Baizhanov 2025). Therefore, we will explore potential benefits and drawbacks, how this approach aligns with global trends and national policies, and the essential conditions required for its practical effectiveness. The core idea is to maintain the strengths of the current system while thoughtfully considering options that improve the UNT's relevance and practicality in today's academic and economic environment.

Alignment with Policy and Global Trends: The reform is well-timed with Kazakhstan's strategic vision of becoming a regional hub for higher education and innovation. The Government's Plan for Higher Education Development (2023–2029) says that the way tests are given needs to be changed. Our proposal directly responds to this call by adding new elements, such as digital literacy and basic AI knowledge, as well as internationalization, such as English academic communication, to the admissions process [12]. Kazakhstan is serious about raising educational standards to global levels by making substantial modifications to the UNT. Comparing countries highlights the significance of this move. Many countries are modifying their tests or adopting new methods to assess skills that are not academic in nature. For instance, Singapore's "Thinking Schools, Learning Nation" reforms place a strong emphasis on critical and creative thinking in tests and lessons [20]. China's recent changes to the Gaokao tests added open-ended questions to see how well students could use what they had learned. Also, the OECD's Education 2030 project and the World Economic Forum's education reports show that creativity, collaboration, and digital fluency are important skills for applicants to have (10; 21). The next-generation UNT is specifically designed to incorporate these skills, ensuring that Kazakhstan's assessment system aligns with 21st-century educational models.

The UNT also adapts to Kazakhstan's higher education by adding an English for Academic Purposes module. This is because an increasing number of programs are being taught in English, making the country more bilingual or trilingual. This resonates with international branch campuses in Kazakhstan, which expect admitted applicants to have specific English academic skills. In brief, the UNT reform will fill in the gap shown by the Heriot-Watt Aktobe case. Future high scorers on the UNT will also have demonstrated that they possess the language and thinking skills that top universities seek, thereby eliminating any potential mismatch, as was the case in the instance cited [5].

Additionally, ensuring that the test content aligns with global standards (for example, utilizing SAT essay tasks, GRE analytical writing, or PISA problem-solving frameworks as models) will enhance the UNT's credibility to international partners. Kazakhstan's universities are increasingly recruiting applicants globally; a modern UNT could eventually serve as a domestic selection tool and

internationally as a signal of applicant quality. In the long run, if modules like EAP are validated, they may be recognized in the same way as IELTS, TOEFL, or other standardized benchmarks, thereby raising the profile of Kazakhstan's educational assessments.

Enhancing University Readiness and Academic Success: A core benefit of the reform is expected to be improved university outcomes. By selecting applicants with a well-rounded skill set, universities will likely see incoming cohorts that need less remedial support and can engage more readily with a higher-level curriculum. For example, an applicant who passes the AWRS module has proven ability in constructing an essay or handling research sources, suggesting they will cope better with university writing assignments. Including a critical thinking score in admissions criteria means that admitted applicants are knowledgeable and can independently analyze and solve problems, a predictor of success in rigorous programs. Consequently, we anticipate lower first-year dropout rates and better academic performance among applicants chosen via the new UNT. This addresses a current issue where some scholarship applicants, admitted solely based on UNT subject scores, struggle academically later. The new UNT admissions process is more akin to a diagnostic test for readiness, similar to an early placement test for skills. This also helps colleges and universities customize their first-year classes. For example, if they know how well a class communicates, they might adjust their teaching approach or offer targeted training in the first semester. In the long run, better matching applicants to programs (through appropriate module combinations and a richer admissions dataset) should increase student success and satisfaction because applicants are admitted based on their demonstrated aptitude in pertinent fields rather than just their interest or rote scores.

From a pedagogical perspective, the reform might have positive washback effects on secondary education. Once schools are aware that the UNT will assess writing, inquiry, and creativity, they are likely to include these skills more in their curriculum. We can expect more classroom activities, such as debates, science projects, essay writing, and group problem-solving, rather than merely memorizing test questions. UNT preparation could become a more fulfilling educational experience if it placed more emphasis on essays, experiments, and portfolios, rather than just multiple-choice exams. In an ideal world, this makes learning more fun for applicants and helps them develop their skills throughout high school. For instance, an applicant preparing for a creative thinking task might undertake independent projects or design challenges, which would help them further develop their skills. Over time, the line between curriculum and test prep may blur healthily: what is good teaching (such as critical analysis of texts or scientific reasoning) will also be effective test prep. This synergy contrasts with the current situation, where “test prep” can sometimes be a narrow pursuit separate from holistic education. Thus, the UNT reform can be a lever for educational improvement, not just a measurement tool [22]. It embodies assessment-driven instruction in a constructive direction.

Human Capital Development and Economic Implications: On a macro level, the shift to a competency-based, modern assessment aligns the education system with Kazakhstan's economic needs. A more innovative, adaptable, and proficient workforce in communication and digital skills is better poised to contribute to a knowledge economy. By encouraging BS to develop these attributes (through the necessity of passing skill modules), the UNT reform can indirectly influence the quality of human capital. Over the years, as cohorts of UNT-takers graduate and enter the workforce, we can expect to see more graduates who possess the critical thinking, problem-solving, and lifelong learning skills that employers value highly in the 21st century. This can enhance productivity and innovation in the economy. Research by Hanushek & Woessmann demonstrates that improvements in cognitive skills (as measured by test scores) have a strong association with economic growth – roughly, one standard deviation increase in national test performance can translate to a 2% higher annual GDP growth rate per capita, all else equal [23, p. 112]. While that statistic typically refers to basic skills tests, one can extrapolate that raising higher-order skills among the populace will similarly benefit the economy. Equipping applicants with digital literacy and familiarity with AI through the ADL module supports Kazakhstan's digitalization agenda. It may help spawn local tech talent, entrepreneurs, and innovators, feeding into initiatives like “Digital Kazakhstan.”

Moreover, a more robust and respected UNT can help retain and attract talent. If top Kazakhstani applicants see that local universities (via the UNT) value the same skills as international ones and that by staying, they get to develop those skills, they might be less inclined to seek undergraduate education abroad. Conversely, if the UNT becomes an exam that international applicants (especially in the region) are willing to take because it is competency-based and perhaps offered in multiple languages, Kazakhstan could attract bright applicants from neighboring countries into its universities, boosting the diversity and quality of its applicant body. This is set to be an academic hub in Eurasia. International student applications are already on the rise [24]; a modern admissions system will make Kazakhstan's higher education more accessible and appealing to them. Over time, integrating international benchmarks enables the UNT to be recognized and used in collaboration with other international tests, thereby further enhancing Kazakhstan's higher education profile.

Challenges and Risk Mitigation: Several challenges must be acknowledged. One risk is the potential for widening inequality if the new skills are unevenly distributed due to differences in school quality. Applicants from well-resourced urban schools might initially fare better in, say, Critical Thinking or English modules compared to those from remote schools, which never had such training. This could inadvertently exacerbate rural-urban gaps, contrary to the UNT's original equity mission. To mitigate this, the reform includes free preparation resources (online platforms with practice tests and tutorial videos for new module tasks) and teacher professional development focusing on teaching and assessing these new skills. Special outreach programs could be implemented for rural schools, such as an "AI and Digital Literacy" crash course for teachers in regions with limited access to technology, funded by the Ministry, so that all applicants receive some exposure before the test rolls out widely. Additionally, in the early years of implementation, score interpretations should be made cautiously. Universities and policymakers could examine performance by school type and region, and possibly norm or contextualize scores so that disadvantaged groups are not unfairly screened out. Over time, as teaching catches up, these disparities are expected to diminish. Indeed, the ambition is that introducing these skills in the exam forces the system to support all applicants in acquiring them, thus raising the floor, not just the ceiling.

Another challenge is operational feasibility and cost. High-quality assessment, including essays, speaking tasks, and sophisticated AI systems, is expensive. It will require sustained political and financial commitment. There may be pushback on the costs of new computer labs, training scorers, licensing AI software, and other related expenses. However, as noted earlier, AI helps control long-term costs [25]. Moreover, these costs should be viewed as an investment in educational quality and human capital. The economic returns of better human capital justify spending on better assessment [26]. Kazakhstan might seek partnerships or technical assistance from international organizations (World Bank, OECD, UNESCO) that often support such reforms to ease the burden. The existing World Bank project could extend to fund some pilot implementations of AI scoring or provide expertise in item development. Engaging local tech companies or universities in the development of AI solutions could also help localize costs and build domestic capacity. A phased rollout also spreads costs over the years. Policymakers should be made aware that cutting corners (e.g., failing to conduct proper pilot testing or attempting to implement everything at once) could jeopardize the validity of the exam and erode public trust. In contrast, measured investment and phased introduction significantly improve the chances of success.

A further challenge is cultural acceptance. Applicants, parents, and teachers have been familiar with the current UNT format for almost 20 years; sudden changes can cause anxiety or even resistance. Some might fear that including soft skills makes the exam subjective or unpredictable. Some people may question AI's role because they fear machines will decide their futures. To address this, the changes should incorporate stakeholder involvement and effective change management. People can stay informed and involved through public forums, Q&A sessions, explanatory videos, and updates on pilot results. Public forums, Q&A sessions, explanatory videos, and updates on pilot results can keep people informed and involved. It is important to highlight that the reform is additive (adding opportunities to

show skills) rather than capricious. The Ministry could emphasize success stories from pilot phases, for example, by showcasing a rural applicant excelling in the creative thinking module due to innate talent, thereby demonstrating that the new UNT can reveal hidden strengths beyond rote learning. Regarding AI, explicit assurances about human oversight and fairness must be communicated, along with evidence (perhaps a white paper) demonstrating the testing of AI systems. Over time, as people see the system working – for example, essays being scored and applicants receiving feedback – confidence will build.

Finally, we must consider that high-stakes changes can have unintended consequences. If not carefully monitored, coaching centers might pivot and find ways to "game" the new system (like formulaic essay templates or tricks to beat AI scoring). Therefore, ongoing monitoring and evaluation mechanisms should be established by responsible stakeholders to identify and address any unintended effects or emerging practices that may compromise the reform's intended goals. Continuous research should accompany implementation, such as studies on how applicants prepare for the new modules and whether those preparations align with genuine skill-building or short-term hacks. The feedback loop should enable adjustments to policies and procedures. For example, if the first high-stakes round indicates that one module is not functioning correctly, it could adjust its weight or reset it to zero. The system should be able to change quickly and easily during the first few years of operation.

Expected Results and Future Prospects. Although there will be challenges, the overall direction of the UNT reform will significantly impact how education is delivered in Kazakhstan. The next-gen UNT will broaden the definition of a successful applicant. The next-gen UNT will change what it means to be a successful applicant. A single score will no longer define them; instead, they can show off a portfolio of skills. This lowers the one-dimensional pressure and may make applicants feel more at ease. For example, an applicant who struggles with memorization but excels at reasoning could still gain admission to college by performing well on the CTL module. This more nuanced meritocracy values a broader range of skills and could make the competition feel more fair because it values different skill sets. Getting ready for the UNT could stop being a boring cycle of doing the same things over and over again and become a learning journey instead. This would make high school more interesting and valuable.

For universities and the higher education system, admissions will become more predictive of success. Committees get a richer data set and can make better-informed decisions [13]. We expect remedial classes for basic writing or computer skills at universities to decrease after a few intakes, as incoming applicants already demonstrate those competencies. This frees up university resources to focus on higher-level teaching, allowing professors to start at a more advanced level and thereby raise academic rigor. In a few years, Kazakhstan's bachelor's graduates may be stronger in writing and critical thinking than they are now, which could also improve outcomes at the graduate education level and enhance research quality.

The UNT reform introduces a culture of continuous assessment and improvement at the system level. With richer data from the exam, the Ministry can pinpoint weaknesses in the school system (e.g., if national results show low scores in creative thinking, that indicates a need to infuse creativity into school pedagogy). The exam thus becomes a tool for feedback and accountability in a positive sense, not to punish schools, but to provide guidance and support. Over time, trends such as rising digital literacy scores can demonstrate that programs like the new ICT curriculum are effective. Kazakhstan could utilize UNT data to compile an annual report on "National Readiness Indicators," thereby contributing to the policy discussion. Kazakhstan can compare its applicants' performance more fairly on a global scale by making the UNT's content and processes more similar to those of other countries (for example, by utilizing IRT humbly and AI, such as ETS or Cambridge Assessment mesh). For example, if its critical thinking module is modeled after known frameworks, it can interpret how its applicants fare on a world stage, which is helpful for policy benchmarking.

The reform also carries symbolic weight: it positions Kazakhstan as an innovator in assessment in the region. Few countries have attempted to integrate as many progressive elements (modularity, multiple languages, AI, creative tasks) in a single national exam at the undergraduate level. If it works,

Kazakhstan's UNT could become a model for other countries, thereby improving the country's reputation in the field of international education. This could lead to opportunities for Kazakhstan to collaborate with others or participate in pilot global programs.

Kazakhstan has a strong base for fairness and standardization in admissions thanks to UNT's successes. The detailed reform suggested here builds on that and aims to make the exam more relevant to current educational needs and Kazakhstan's strategic goals. It is an ambitious reform – touching content, format, technology, and policy – but research and international experience have carefully justified each of its elements. With phased implementation, stakeholder buy-in, and careful monitoring, the next-generation UNT can be realized with manageable risk. The result will be a high-stakes assessment that not only selects university applicants but also drives secondary education towards broader skills, provides valuable feedback for continuous improvement, and contributes to the development of a skilled and innovative workforce. By doing this, Kazakhstan's UNT will have transitioned from a successful gatekeeper for university access to a driver of high-quality education and a key contributor to the development of human capital.

Conclusion. The UNT has made it fair and open for people to get into college. However, the UNT needs to move beyond simply recalling information and toward a more comprehensive assessment model, as the needs of higher education and the economy are changing rapidly. The proposed reform would establish a modular, competency-based framework that integrates subject knowledge with essential 21st-century skills, including critical thinking, academic writing, communication, digital literacy, and creativity. These changes retain UNT's best features while enhancing its ability to identify applicants who are ready for college and their future careers. These suggestions are offered as a progressive, adaptable choice that can be modified based on stakeholder feedback and assessed for equity. By presenting it this way, the UNT can uphold its commitment to equity while playing a role in national growth as an effective tool for developing human capital [6].

The reform uses new psychometric methods and AI technologies to ensure that tests are fair, thorough, and quick. A phased implementation approach, which includes a pilot, a scale-up, and a full deployment, is a smart way to make the change. The reform is more than just a technical change; it demonstrates that Kazakhstan aims to cultivate a population that is prepared for the future, creative, and skilled. The next-generation UNT promises to be a fair selection tool and a means to improve the entire education system and develop human capital in Kazakhstan, provided it is carefully planned and all stakeholders are involved.

Список использованных источников:

1. *Organisation for Economic Co-operation and Development (OECD); The World Bank. Reviews of National Policies for Education: Higher Education in Kazakhstan. Paris: OECD; Washington, D.C.: The World Bank, 2007. 226 p. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2007/06/reviews-of-national-policies-for-education-higher-education-in-kazakhstan-2007_g1gh7f86/9789264033177-en.pdf (Accessed 20 October 2025).*

2. *Amankulova Z. How rurality affects access to higher education for applicants in Kazakhstan //International Journal of Multidisciplinary Perspectives in Higher Education. – 2018. – Vol. 3, No. 1. – P. 4–18.*

3. *World Bank. Charting the Future of Assessment: A Strategic Review of the Unified National Testing in Kazakhstan. – World Bank Group, Washington DC, 2023. – pp. 30-31, 42-50, 84-89.*

4. *Министерство науки и высшего образования Республики Казахстан. Правила проведения Единого национального тестирования [Электронный ресурс]. – Астана, 2023. – Режим доступа: https://adilet.zan.kz/eng/docs/V14H0009940?utm_source*

5. *Уразалина А. Более ста студентов вуза в Актобе отчислены накануне учебного года [Электронный ресурс] // Liter.kz. – 27 авг. 2024 г. – Режим доступа: https://liter.kz/bolee-sta-studentov-vuza-v-aktobe-otchisleny-nakanune-uchebnogo-goda-1724735401/*

6. *World Bank. Equitable Human Capital Development in Kazakhstan : A Policy Overview. Washington, D.C.: World Bank Group, 2021. – Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099083002082350371 (Accessed 20 October 2025).*

7. *Organisation for Economic Co-operation and Development (OECD). What Students Learn Matters:Towards a 21st Century Curriculum. Paris: OECD Publishing, 2020. – Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/11/what-students-learn-matters_555a22ec/d86d4d9a-en.pdf (Accessed 20 October 2025).*

8. Байжанов Н., The Next-Generation UNT Vision Paper [неопубликованная рукопись]. Алматы,2025.

9. Zharkynbekova Sh., Issenova F. "Academic writing" in the system of higher education disciplines in Kazakhstan: problems of theoretical justification and practical approbation. – Bulletin of L.N. Gumilyov Eurasian National University. Philology Series, 2021, vol. 135, no. 2, pp. 100–109. – DOI: https://doi.org/10.32523/2616-678X-2021-135-2-100-109

10. Organisation for Economic Co-operation and Development (OECD). PISA 2022 Creative Thinking Assessment Framework. – OECD Publishing, Paris, 2023.

11. Organisation for Economic Co-operation and Development (OECD). Education Policy Outlook 2023:Trends and Policy Priorities. – OECD Publishing, Paris, 2023.

12. Government of the Republic of Kazakhstan. Concept for the Development of the Higher Education and Science System of the Republic of Kazakhstan for 2023–2029. – Astana, 2023. – pp. 3, 5–7, 17.

13. Liu O. L., Frankel L., Roohr K. C. Assessing critical thinking in higher education: Current state and directions for next-generation assessment // ETS Research Report Series. – 2014. – Vol. 2014, No. 1. – P. 1–23.

14. Watson, G., Glaser, E. M. Watson-Glaser II Critical Thinking Appraisal: Technical Manual and User's Guide. Bloomington, MN: Pearson / NCS Pearson, 2009.

15. Liu O. L., Frankel L., Roohr K. C. Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment // ETS Research Report Series. – 2014. – Vol. 2014, No. 1. – P. 1–23. – DOI: 10.1002/ets2.12009.

16. Organisation for Economic Co-operation and Development (OECD). 21st-Century Readers: Developing Literacy Skills in a Digital World. – OECD Publishing, Paris, 2021.

17. Organisation for Economic Co-operation and Development (OECD). Strengthening National Examinations in Kazakhstan to Achieve National Goals. – Education Policy Perspectives, No. 24. – OECD Publishing, Paris, 2020. – DOI:10.1787/0bf8662b-en.

18. Zhao, Y. What Works May Hurt: Side Effects in Education. – New York: Teachers College Press, 2018.– 176 p.

19. Zhao, Y. An Education Crisis Is a Terrible Thing to Waste: How Radical Changes Can Spark Student Excitement and Success. – New York: Teachers College Press, 2019. – 160 p.

20. Tan, J., & Ng, P. T. (eds.). Thinking Schools, Learning Nation: Contemporary Issues and Challenges. – Singapore: Pearson, 2008. – 278 p.

21. World Economic Forum. New Vision for Education: Fostering Social and Emotional Learning through Technology. – World Economic Forum (WEF) Report, Geneva, 2016.

22. Kellaghan, T., & Greaney, V. (2019). Public examinations examined. World Bank. https://doi.org/10.1596/978-1-4648-1418-1

23. Hanushek E. A., Woessmann L. The knowledge capital of nations: Education and the economics of growth. – MIT press, 2023.

24. Temirgaliyeva, A. 23 Branches of Foreign Universities Operate in Kazakhstan Today [Electronic resource] // Kazinform International News Agency. – June 5, 2025. – Mode of access: https://qazinform.com/news/23-branches-of-foreign-universities-operate-in-kazakhstan-today-07d709 (Accessed 20 October 2025).

25. Topol B., Olson J., Roeber E. The cost of new higher quality assessments: A comprehensive analysis of the potential costs for future state assessments //Assessment Solutions Group. Stanford, CA: Stanford Center for Opportunity Policy in Education. – 2010.

26. Hanushek E.A., Woessmann L. The knowledge capital of nations: Education and the economics of growth. – MIT press, 2023.

References:

1. Organisation for Economic Co-operation and Development (OECD); The World Bank. Reviews of National Policies for Education: Higher Education in Kazakhstan. Paris: OECD; Washington, D.C.: The World Bank, 2007. 226 p. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2007/06/reviews-of-national-policies-for-education-higher-education-in-kazakhstan-2007_g1gh7f86/9789264033177-en.pdf (Accessed 20 October 2025).

2. Amankulova Z. How rurality affects access to higher education for applicants in Kazakhstan // International Journal of Multidisciplinary Perspectives in Higher Education. – 2018. – Vol. 3, No. 1. – P. 4–18.

3. World Bank. Charting the Future of Assessment: A Strategic Review of the Unified National Testing in Kazakhstan. – World Bank Group, Washington DC, 2023. – pp. 30-31, 42-50, 84-89.

4. Ministry of Science and Higher Education of the Republic of Kazakhstan. Rules for Conducting the Unified National Testing [Electronic resource]. – Adilet.zan.kz (legal portal), 2023. – Available at: https://adilet.zan.kz/eng/docs/V14H0009940?utm_source (Accessed 20 October 2025).

5. Urazalina, A. More than one hundred university students in Aktobe were expelled on the eve of the academic year. – Liter.kz, August 27, 2024. – Available at: https://liter.kz/bolee-sta-studentov-vuza-v-aktobe-otchisleny-nakanune-uchebnogo-goda-1724735401/ (Accessed 20 October 2025).

6. World Bank. Equitable Human Capital Development in Kazakhstan : A Policy Overview. Washington, D.C.: World Bank Group, 2021. – Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099083002082350371 (Accessed 20 October 2025).

7. Organisation for Economic Co-operation and Development (OECD). What Students Learn Matters: Towards a 21st Century Curriculum. Paris: OECD Publishing, 2020. – Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/11/what-students-learn-matters_555a22ec/d86d4d9a-en.pdf (Accessed 20 October 2025).

8. Baizhanov N., The Next-Generation UNT Vision Paper [unpublished work рукопись]. Almaty, 2025.

9. Zharkynbekova Sh., Issenova F. "Academic writing" in the system of higher education disciplines in Kazakhstan: problems of theoretical justification and practical approbation. – Bulletin of L.N. Gumilyov Eurasian National University. Philology Series, 2021, vol. 135, no. 2, pp. 100–109. – DOI: https://doi.org/10.32523/2616-678X-2021-135-2-100-109

10. Organisation for Economic Co-operation and Development (OECD). PISA 2022 Creative Thinking Assessment Framework. – OECD Publishing, Paris, 2023.

11. Organisation for Economic Co-operation and Development (OECD). Education Policy Outlook 2023:Trends and Policy Priorities. – OECD Publishing, Paris, 2023.

12. Government of the Republic of Kazakhstan. Concept for the Development of the Higher Education and Science System of the Republic of Kazakhstan for 2023–2029. – Astana, 2023. – pp. 3, 5–7, 17.

13. Liu O. L., Frankel L., Roohr K. C. Assessing critical thinking in higher education: Current state and directions for next-generation assessment // ETS Research Report Series. – 2014. – Vol. 2014, No. 1. – P. 1–23.

14. Watson, G., Glaser, E. M. Watson-Glaser II Critical Thinking Appraisal: Technical Manual and User's Guide. Bloomington, MN: Pearson / NCS Pearson, 2009.

15. Liu O. L., Frankel L., Roohr K. C. Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment // ETS Research Report Series. – 2014. – Vol. 2014, No. 1. – P. 1–23. – DOI: 10.1002/ets2.12009.

16. Organisation for Economic Co-operation and Development (OECD). 21st-Century Readers: Developing Literacy Skills in a Digital World. – OECD Publishing, Paris, 2021.

17. Organisation for Economic Co-operation and Development (OECD). Strengthening National Examinations in Kazakhstan to Achieve National Goals. – Education Policy Perspectives, No. 24. – OECD Publishing, Paris, 2020. – DOI:10.1787/0bf8662b-en.

18. Zhao, Y. What Works May Hurt: Side Effects in Education. – New York: Teachers College Press, 2018.– 176 p.

19. Zhao, Y. An Education Crisis Is a Terrible Thing to Waste: How Radical Changes Can Spark Student Excitement and Success. – New York: Teachers College Press, 2019. – 160 p.

20. Tan, J., & Ng, P. T. (eds.). Thinking Schools, Learning Nation: Contemporary Issues and Challenges. –Singapore: Pearson, 2008. – 278 p.

21. World Economic Forum. New Vision for Education: Fostering Social and Emotional Learning through Technology. – World Economic Forum (WEF) Report, Geneva, 2016.

22. Kellaghan, T., & Greaney, V. (2019). Public examinations examined. World Bank. https://doi.org/10.1596/978-1-4648-1418-1

23. Hanushek E. A., Woessmann L. The knowledge capital of nations: Education and the economics of growth. – MIT press, 2023.

24. Temirgaliyeva, A. 23 Branches of Foreign Universities Operate in Kazakhstan Today [Electronic resource] // Kazinform International News Agency. – June 5, 2025. – Mode of access: https://qazinform.com/news/23-branches-of-foreign-universities-operate-in-kazakhstan-today-07d709 (Accessed 20 October 2025).

25. Topol B., Olson J., Roeber E. The cost of new higher quality assessments: A comprehensive analysis of the potential costs for future state assessments //Assessment Solutions Group. Stanford, CA: Stanford Center for Opportunity Policy in Education. – 2010.

26. Hanushek E. A., Woessmann L. The knowledge capital of nations: Education and the economics of growth. – MIT press, 2023.