



ULTTYQ
TESTILEÝ
ORTALYGY

ISSN: 3105-4161 (Online)

ПЕДАГОГИКАЛЫҚ ӨЛШЕМДЕР

ҒЫЛЫМИ-ПРАКТИКАЛЫҚ ЖУРНАЛЫ

№3/2026



ПЕДАГОГИЧЕСКИЕ
ИЗМЕРЕНИЯ

НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

PEDAGOGICAL
MEASUREMENTS

JOURNAL OF RESEARCH AND PRACTICE

Қазақстан Республикасының Ғылым және жоғары білім министрлігі «Ұлттық тестілеу орталығы»
шаруашылық жүргізу құқығындағы республикалық мемлекеттік кәсіпорны

«ПЕДАГОГИКАЛЫҚ ӨЛШЕМДЕР» ҒЫЛЫМИ-ПРАКТИКАЛЫҚ ЖУРНАЛЫ

Республиканское государственное предприятие на праве хозяйственного ведения «Национальный центр
тестирования» Министерства науки и высшего образования Республики Казахстан

НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ «ПЕДАГОГИЧЕСКИЕ ИЗМЕРЕНИЯ»

Republican state enterprise under the right of economic management «National Testing Center»
of the Ministry of Science and Higher Education of the Republic of Kazakhstan

JOURNAL OF RESEARCH AND PRACTICE «PEDAGOGICAL MEASUREMENTS»

№3(7) / 2026

*Жылына 4 рет шығарылады
2025 жылдан бастап шығады*

Астана
2026

ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК

«ПЕДАГОГИКАЛЫҚ ӨЛШЕМДЕР»
ғылыми-практикалық журналы*Бас редакторы:***Абдрасилов Б.С.**

ҚР ҒЖБМ Ұлттық тестілеу орталығының директоры, ҚР ҰҒА академигі

*Жауапты редактор:***Шинетова Л.Е.***Редакция алқасы:*

Matthew Courtney – PhD, Назарбаев Университет

Карданова Е. Ю. – ф.-м.ғ.к., доцент, «Жоғары экономика мектебі» Ғылыми-зерттеу университеті, Ресей

Иванова А.Е. – Білім беру ғылымдарының кандидаты, доцент, «Жоғары экономика мектебі» Ғылыми-зерттеу университеті, Ресей

Зулпыхар Ж.Е. – п.ғ.к., Л.Н. Гумилев атындағы Еуразия ұлттық университеті

Калкеева К.Р. – п.ғ.к., профессор, Л.Н. Гумилев атындағы Еуразия ұлттық университеті

Кошербаева А.Н. – п.ғ.д., профессор, Абай атындағы Қазақ ұлттық педагогикалық университеті

Танирбергенова А.Ш. – п.ғ.к., Ы.Алтынсарин атындағы Ұлттық білім академиясы

Абаева Г.А. – п.ғ.к., доцент, Абай атындағы Қазақ ұлттық педагогикалық университеті

Абсатова М.А. – п.ғ.к., профессор, Абай атындағы Қазақ ұлттық педагогикалық университеті

Акижанова Д.М. – PhD, Л.Н. Гумилев атындағы Еуразия ұлттық университеті

Бегимбетова Г.А. – Universitas Negeri Yogyakarta, Индонезия

Дидарбекова Н.А. – ф.ғ.к., Ұлттық тестілеу орталығы

Temur Nezir – PhD, профессор, Гази университеті, Түркия

*Техникалық редактор:***Бейсенова Ж.Ж.****«Педагогикалық өлшемдер» ғылыми-практикалық журналы. – 2026. – 7 т. – 229 б.**Редакцияның мекенжайы: 010000, Қазақстан Республикасы, Астана қ., Жеңіс даңғылы, 60, тел. 8(7172)799-889, і.н.157, e-mail: naukauto24@gmail.com, сайт: <https://testcenter-zhurnal.kz/>.

Қазақстан Республикасы Мәдениет және ақпарат министрлігімен тіркелген. №KZ54VPY00116912 есепке қою туралы куәлігі.

МАЗМҰНЫ / СОДЕРЖАНИЕ / CONTENT

БІЛІМ БЕРУ НӘТИЖЕЛЕРІН БАҒАЛАУДЫҢ ЗАМАНАУИ ТӘСІЛДЕРІ ЖӘНЕ БІЛІМ БЕРУ САПАСЫН БАСҚАРУ / СОВРЕМЕННЫЕ ПОДХОДЫ К ОЦЕНКЕ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ И УПРАВЛЕНИЕ КАЧЕСТВОМ ОБРАЗОВАНИЯ / MODERN APPROACHES TO EDUCATIONAL OUTCOMES ASSESSMENT AND EDUCATION QUALITY MANAGEMENT

M. Khassenova, M. Smagulov, M. Erkibayeva

DIGITAL SCREENING AND MONITORING OF LEARNING DYNAMICS AMONG STUDENTS WITH SPECIAL EDUCATIONAL NEEDS IN NATURAL SCIENCE EDUCATION..... 5

А.Б. Ертілеуова, Р. Желвис

БОЛАШАҚ ПЕДАГОГТАРДЫҢ КӘСІБИ ДАЯРЛЫҒЫНДА МЕКТЕП ТӨП-МЕНЕДЖЕРІНІҢ САУАТТЫЛЫҒЫН ҚАЛЫПТАСТЫРУ..... 19

Ә.Қ. Құлмағанбет, Д.С. Оспанов, Т.С. Ерсұлтанова

ПЕДАГОГИКА ЖӘНЕ ПСИХОЛОГИЯ БОЙЫНША ТЕСТІЛЕУДІ ПСИХОМЕТРИЯЛЫҚ ТАЛДАУ: ТЕХНИКАЛЫҚ ЖӘНЕ КӘСІПТІК БІЛІМ БЕРУ ЖҮЙЕСІНДЕГІ БАҒАЛАУ САПАСЫНЫҢ КЕПІЛІ..... 31

N. Issina, A. Yerzhumanova, G. Baideldinova

GENERATION Z STUDENTS' PREFERENCES IN LEARNING EFL IN KAZAKHSTAN: AN ONLINE SURVEY STUDY..... 45

S. Kastayev, A. Tursynbayeva, N. Sartayeva, S. Mambetaliyev, A. Bostanova

DIAGNOSTICS OF THE FORMATION OF ORGANIZATIONAL SKILLS OF FUTURE TEACHERS..... 61

У.З. Ешимова, Б.Б. Махмұтов, Ж.О. Оразымбетова, З.Н. Маханбетова

ПЕДАГОГИКАЛЫҚ КОЛЛЕДЖ СТУДЕНТТЕРІНІҢ ПЕДАГОГИКАЛЫҚ ПРАКТИКАСЫНДА КӘСІБИ ЖӘНЕ ЖҰМСАҚ DAҒДЫЛАРЫН БАҒАЛАУДЫҢ ПЕДАГОГИКАЛЫҚ ШАРТТАРЫ МЕН ӨЛШЕМДЕРІ..... 75

Э.Ж. Шамшинова

ЛИНГВОДИДАКТИЧЕСКИЕ ОСНОВЫ ФОРМИРОВАНИЯ ФРАЗОВОЙ РЕЧИ ПРИ ЗРИТЕЛЬНОЙ ДЕПРИВАЦИИ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ НАУЧНЫХ ПОДХОДОВ И МОДЕЛЬ КРИТЕРИАЛЬНОГО ОЦЕНИВАНИЯ..... 91

Ж.Ж. Бейсенова, Л.Е. Шинетова, Г.К. Жабаева, С.Ж. Анафияева

ДИФФЕРЕНЦИАЛЬНОЕ ФУНКЦИОНИРОВАНИЕ ТЕСТОВЫХ ЗАДАНИЙ В ТЕСТИРОВАНИИ ПЕДАГОГОВ НА ОСНОВЕ МОДЕЛИ РАША..... 107

L. Rizayeva, A. Zhunissova, L. Nurseitova, K. Turalbaeva, L. Seisekenova

DIAGNOSTICS OF THE FORMATION OF RESEARCH SKILLS OF PRIMARY SCHOOL CHILDREN IN MATHEMATICS LESSONS..... 119

Р.С. Қалиева, Г.Ш. Ерназарова, М.Н. Қойшыбаев, К.С. Кенжебаева

БАСТАУЫШ СЫНЫП ОҚУШЫЛАРЫНЫҢ КІТАП ОҚУҒА ҚЫЗЫҒУШЫЛЫҒЫН АРТТЫРУДА ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАРДЫ ҚОЛДАНУ: МҰҒАЛІМДЕРДІҢ ТӘЖІРИБЕСІ МЕН БАҒАЛАУЫ..... 137

Б.Б. Топоева-Ставинская, Г.С. Ерсұлтанова

КОНЦЕПТУАЛЬНО-МЕТОДИЧЕСКИЕ ОСНОВЫ РАЗРАБОТКИ УРОВНЯ С2 В НАЦИОНАЛЬНЫХ СИСТЕМАХ ЯЗЫКОВОГО ТЕСТИРОВАНИЯ «КЫРГЫЗТЕСТ» И КАЗТЕСТ: СРАВНИТЕЛЬНАЯ МОДЕЛЬ ПЯТИ- И ЧЕТЫРЕХКОМПОНЕНТНОЙ АРХИТЕКТУРЫ..... 154

Ш.Б. Алтыбаева, А.Н. Қуанышбекова, Ж. Абдулла, А.Н. Есимбекова

ОЦЕНКА КОГНИТИВНОЙ СЛОЖНОСТИ ЗАДАНИЙ ПО ХИМИИ: ДАННЫЕ ЕНТ КАЗАХСТАНА..... 167

**ЖАСАНДЫ ИНТЕЛЛЕКТ ЖӘНЕ БІЛІМ БЕРУДІ БАҒАЛАУДАҒЫ ИННОВАЦИЯ /
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ИННОВАЦИИ В ОЦЕНИВАНИИ ОБРАЗОВАНИЯ
/ ARTIFICIAL INTELLIGENCE AND INNOVATIONS IN EDUCATIONAL ASSESSMENT**

A. Кушубек, Б. Шарипбеков, Б. Рахимов, С. Сантеева, А. Шайханова

IT-ИНФРАҚҰРЫЛЫМДАРДЫҢ АҚПАРАТТЫҚ ҚАУІПСІЗДІГІН БАСҚАРУДА
ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕГІ ЛОГ ТАЛДАУ ЖӘНЕ
АВТОМАТТАНДЫРЫЛҒАН ӨРЕКЕТ ЕТУ АРХИТЕКТУРАЛАРЫНА ШОЛУ..... 183

A. Zhorabekova, M. Duisenova, M. Kurmasheva, A. Yesnazar

MEASUREMENT INVARIANCE OF AN AI-GAMIFIED ADAPTIVE LEARNING
ACCEPTANCE SCALE ACROSS ENGLISH PROFICIENCY AND YEAR OF STUDY..... 197

M.E. Ospanaly, A.K. Kalieva, N. Temur

EVALUATING AI RESPONSES IN LITERATURE TEACHING: THE LITERARY INTERIOR
IN KAZAKH PROSE..... 208

A.U. Serikbayeva, D.M. Akizhanova

CAN GENERATIVE AI REPLACE HUMAN ESSAY RATERS? A PSYCHOMETRIC
COMPARISON OF GPT-5.2, GEMINI, AND TEACHER RATINGS..... 218

**БІЛІМ БЕРУ НӘТИЖЕЛЕРІН БАҒАЛАУДЫҢ ЗАМАНАУИ ТӘСІЛДЕРІ ЖӘНЕ
БІЛІМ БЕРУ САПАСЫН БАСҚАРУ**

**СОВРЕМЕННЫЕ ПОДХОДЫ К ОЦЕНКЕ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ И
УПРАВЛЕНИЕ КАЧЕСТВОМ ОБРАЗОВАНИЯ**

**MODERN APPROACHES TO EDUCATIONAL OUTCOMES ASSESSMENT AND
EDUCATION QUALITY MANAGEMENT**

M. Khasanova¹, M. Smagulov^{2*}, M. Erkibayeva³^{1,2,3}A. Margulan Pavlodar Pedagogical University, Pavlodar, Republic of Kazakhstan

*e-mail: meirmansmagulov@gmail.com

¹ORCID 0000-0001-7672-7407, ²ORCID 0009-0005-2939-9275, ³ORCID 0000-0002-6586-2125

DIGITAL SCREENING AND MONITORING OF LEARNING DYNAMICS AMONG STUDENTS WITH SPECIAL EDUCATIONAL NEEDS IN NATURAL SCIENCE EDUCATION

This study tackles the need for objective assessment of learning dynamics among students with special educational needs (SEN) in inclusive education in Kazakhstan. Conventional summative and criteria-based assessment mainly records outcomes. They often do not capture individual learning pace, cognitive demand, recurring error patterns, or the degree of independence demonstrated during task completion. These limitations are especially important in natural science subjects, where learners must work with abstract concepts, visual representations, formulas, and causal relationships. The purpose of the study was to develop and empirically test a model of digital screening and continuous pedagogical monitoring in inclusive sixth-grade natural science classes. The study involved 34 students with SEN and was conducted from January to May 2026. Google Classroom and electronic progress logs were used to record digital learning indicators, including task scores, completion time, number of attempts, regularity of participation, and use of learning resources. The data were evaluated using descriptive statistics, a paired-samples Student's t-test, and Pearson correlation analysis. The mean achievement score increased from 61.8% at pre-test to 76.9% at post-test, while variability decreased. The difference was statistically significant, $t(33) = 6.18$, $p < .001$. Digital task scores were strongly associated with final achievement ($r = .74$, $p < .001$), and regular task completion also showed a positive relationship with performance ($r = .66$, $p < .001$). The findings show that digital screening supports early identification of learning difficulties, timely feedback, and more flexible adjustment of individual educational trajectories in inclusive natural science education.

Keywords: inclusive education, special educational needs, digital screening, pedagogical monitoring, learning analytics, natural science education.

Introduction

Today, digital transformation and the development of an inclusive society in Kazakhstan require a new approach to learning: high-quality knowledge needs to be accessible to every child. This task is most difficult to implement when teaching natural science subjects to students with special educational needs (SES). Such subjects require developed abstract and logical thinking, as well as the ability to build causal relationships, which often causes cognitive difficulties in children.

Although the updated content of education and criteria-based assessment (summative assessment for a unit and for a term) are actively applied in the republic's schools, they still record only the final result. The learning process itself falls out of the teacher's field of vision. Traditional control does not allow you to notice in time how fast the child is learning the material, how he copes with the cognitive strain, where exactly he makes mistakes, and how independent he is in his work.

The paradox of the modern school is that the mass introduction of virtual platforms (LMS) has been reduced to formal accounting of academic performance. The powerful analytical potential of IT tools is practically not used for pedagogical diagnostics and flexible configuration of the educational process. As a result, there is an obvious gap between technologic capabilities and real inclusive practices [1].

This conclusion is supported by recent research in Learning Analytics, which indicates that educational platforms collect substantial volumes of digital data. However, this data is rarely used to identify learning difficulties faced by students with disabilities and special educational needs in a

timely manner. Instead of facilitating pedagogical decision-making, the analytical capabilities of learning management systems (LMSs) tend to focus more on administrative monitoring and the recording of learning outcomes [2, 3].

Creating evidence-based digital screening and continuous assessment models helps solve this problem. By tracking the student's "digital footprint", the teacher can quickly interpret the data and instantly adjust the individual educational route [4]. This approach alters evaluation from a statement of fact into a living, formative tool. This is critically important when reviewing complex natural sciences. In this context, the integrative course "Natural Science" in the 6th grade becomes a key propaedeutic basis that prepares students with SEN for further development of isolated disciplines-chemistry, physics, and biology.

Recent research in Educational Data Mining and Learning Analytics has shown that analyzing students' digital footprints allows us to identify consistent patterns of errors, predict academic risks, and effectively adapt the educational process to meet individual students' needs in a timely manner [3].

Research relevance. The contradictions outlined above show that equal access to complex subject knowledge isn't merely a declarative goal and a major challenge for modern schools in Kazakhstan. Natural science subjects are particularly demanding for students with SEN because they involve abstract concepts, formulas, visual models, and complex causal relationships. If the dynamics of a learner's cognitive difficulties are not identified at an early stage, subsequent study of separate disciplines such as chemistry, physics, and biology may become substantially more difficult.

A major hindrance to effective inclusion is the "blind spot" in conventional assessment. Summative assessment for a unit and for a term records whether a learner has achieved the expected result, but it provides limited information about the learning process itself. For students with SEN, teachers need to understand how quickly fatigue develops, how rapidly didactic units are mastered, where reasoning breaks down, when assistance is required, and when a learner can work independently.

This creates an urgent need to move from slow control to adaptive digital technologies. Automated digital screening can continuously record a learner's digital footprint during task performance, permitting teachers not only to identify existing gaps and also to anticipate emerging difficulties and adjust an individual educational trajectory before persistent underachievement or learning stress develops.

Analysis of the level of research of the problem. Issues of inclusive education and the use of digital technologies in pedagogy are in the focus of attention of both domestic and foreign researchers.

The theoretical and methodological foundations of the development of inclusive education in Kazakhstan, as well as issues of personnel training and psychological and pedagogical support for students with SEN, are described in detail in the works of Kazakhstani scientists. In particular, the works of Z. A. Movkebayeva, G. A. Abayeva, and B. A. Dyusenbayeva [5, 6, 7] substantiate the concept of individualization of the educational process, reveal approaches to the decomposition of professional competencies of special teachers, and point out the critical importance of creating an modifiable educational environment.

Monitoring aspects and problems of implementing digital tools in Kazakhstani schools is considered in the research of K. M. Berkimbayev[8], who points out the demand for a comprehensive digital transformation of student knowledge control.

In the near abroad, the problems of special pedagogy and differentiation of training are actively developed by N. M. Nazarova and O. I. Kukushkina [9]. Researchers of the Institute of Correctional Pedagogy substantiate the concept of "scaffolding tools", proving that computer technologies can act as cognitive supports for children with mental and speech disorders.

The next stage in the development of this research area involved the application of Learning Analytics technologies focused on analysing digital data generated during the educational process to personalise learning. In their review, C. Baek and S. J. Aguilar analysed publications on Learning

Analytics from 2011 to 2020 concerning students with disabilities and special educational needs. The authors found that only about one-third of the studies directly focused on this group of learners, whereas most addressed the early identification of learning difficulties, the prediction of educational risks, the accessibility of digital educational technologies, and support for pedagogical decision-making. The researchers emphasise the need to develop specialised Learning Analytics models that account for the specific characteristics and needs of students with special educational needs and facilitate timely pedagogical intervention [10].

In foreign countries, there is an exponential growth in the number of publications devoted to the use of intelligent adaptive systems (Intelligent Tutoring Systems) in inclusion. According to a systematic review by T. A. Mikropoulos and G. Iatraki, digital technologies are an effective means of supporting the learning of students with special educational needs in the study of natural science disciplines. Their use contributes to improving academic results, increasing student motivation and involvement, and creating conditions for objective monitoring of their educational achievements [11].

These findings are further supported by the systematic review and meta-analysis conducted by S. E. Wilson, W. J. Therrien, and colleagues, which examined the effectiveness of Inquiry-Based Science Instruction for students with disabilities and special educational needs in science education. Based on an analysis of 26 empirical studies, the authors found that the use of inquiry-based teaching methods, combined with pedagogical support, content adaptation, and cognitive scaffolding, contributes to statistically significant improvements in students' subject knowledge and inquiry skills. At the same time, the authors emphasize that sustainable educational outcomes can be achieved only through systematic monitoring of learning progress and timely adjustment of pedagogical support. This confirms the relevance of developing digital tools for screening and monitoring the learning progress of students with special educational needs [12].

A systematic review by García Terceño and Greca analyzed 27 studies on science education for students with special educational needs. The authors found that the most effective approaches were learner-centered and combined inquiry-based learning, hands-on activities, content adaptation, and explicit instruction. They further highlighted the limited evidence base and the need for evidence-based models for assessing achievement and monitoring educational progress among students with SEN [13]. Wade et al. presented evidence-based recommendations for digital monitoring of Individualized Education Program (IEP) goals in distance learning. Their model included electronic forms, mobile applications, videoconferencing platforms, and online services for collecting and analyzing achievement data and supporting collaboration between teachers and families [14].

The evidence base on Learning Progress Monitoring (LPM) is further summarized in the meta-analysis by Fuchs, Radkowsch, and Sommerhoff [15]. Their findings are particularly relevant to the present study because they demonstrate that regular collection and analysis of achievement data can support timely pedagogical assistance and alteration of the learning process.

The literature review reveals two central contradictions. At the theoretical and methodological level, the recognized need to individualize the education of students with SEN is not yet supported by clearly validated digital markers capable of measuring learning microdynamics in real time. At the practice-oriented level, the rapid introduction of instructional platforms in Kazakhstan is not matched by a sufficient number of adapted digital screening methods for natural science subjects that take into account individual educational needs.

The identified research gap is that most digital assessment systems are designed for typically developing learners. In contrast, foreign adaptive systems are not sufficiently aligned with Kazakhstan's curricula and criteria-based assessment framework.

The object of the study is the process of teaching natural science subjects to students with special educational needs in inclusive education. The subject of the study is a pedagogical model of digital screening and monitoring of learning dynamics among students with SEN in natural science education.

The study attempts to theoretically substantiate, develop, and experimentally test the effectiveness of the digital screening and persistent monitoring model designed for optimizing the teaching of natural science subjects to school students with SEN.

The objectives were: (1) to identify cognitive barriers experienced by students with SEN when mastering natural science concepts; (2) to determine digital markers reflecting learning dynamics, including completion time, error patterns, attempts, and use of hints; (3) to develop and implement adapted digital tasks; and (4) to assess the effect of continuous digital monitoring on science achievement and learning dynamics.

The scientific novelty of the study lies in the operationalization of learning dynamics utilizing a set of digital markers and in their embedding into a continuous monitoring model adapted to inclusive sixth-grade natural science education. The practical significance lies in the possibility of using this model for early identification of learning difficulties and timely adjustment of individual educational trajectories.

Research materials and methods

The research intends to develop and test a model for digital screening and monitoring the learning dynamics of students with special educational needs (SEN) when studying subjects of the natural science cycle. The methodological basis of the study is competence-based, system-activity, all-encompassing and data-oriented approaches that consider digital educational technologies as a tool for objective assessment and monitoring of individual educational progress of students.

To achieve this goal, a set of complementary theoretical and empirical research methods was used.

Theoretical methods included analysis, synthesis, comparison, generalization, and systematization of domestic and foreign scientific publications indexed in the Scopus and Web of Science databases on digital assessment, learning analytics, Learning Progress Monitoring, digital screening, and training of students with special educational needs. The analysis of scientific literature allowed us to identify current trends in the use of digital technologies for diagnostics and supervision of students' educational achievements.

The empirical part of the study was conducted in general education schools in Pavlodar, with the participation of sixth-grade students who study the natural science "Natural Science", which is a prerequisite for the disciplines of biology, physics, and chemistry. The study involved 34 students with special educational needs who have the results of psychological, medical, and pedagogical consultation (PMPK) (on which our attention was focused). The study was conducted between January and May 2026.

Digital monitoring was implemented through Google Classroom and Kundelik.kz, supplemented by digital diagnostic tasks and electronic progress logs.

The study recorded indicators of students' learning activity, including the results of performing diagnostic tasks, the number of correct answers, the time required to complete tasks, the frequency of access to digital educational resources, the number of attempts to complete tasks, and the dynamics of learning material development.

The data were examined using IBM SPSS Statistics. Mean values (M), standard deviations (SD), and medians (Me) were calculated to describe the sample. The statistical significance of pre-test and post-test differences was assessed using a paired-samples Student's t-test. Pearson correlation analysis was used to examine relationships between digital monitoring indicators and learning results. Statistical significance was set at $p < .05$.

The reliability of the research results was ensured by the integrated use of qualitative and quantitative methods of analysis, comparison of data from various sources, as well as compliance with uniform conditions for digital diagnostics and observation of students' educational achievements.

The study followed a one-group pre-test–post-test design. The intervention consisted of regular digital diagnostic tasks, automated recording of task-performance indicators, and teacher feedback

based on the collected data. The same assessment framework was applied at baseline and at the final stage to ensure comparability of the results.

Results and discussion

The study involved a group of 34 students with special educational needs. The study aimed to determine the dynamics of the development of the knowledge component when using an electronic educational platform in the learning process.

At the first stage, digital screening was conducted to determine the initial level of sixth-grade students' mastery of the subject "Natural Science," identify individual learning difficulties, and establish a baseline for subsequent monitoring. Design of the pedagogical experiment:

Students with SEN (n = 34)



PRE-TEST (baseline assessment)



DIGITAL MONITORING INTERVENTION

Google Classroom, adapted tasks, feedback, and learning analytics



POST-TEST (final assessment)



STATISTICAL ANALYSIS (paired t-test and correlations)

Analysis of the data presented in Table 1 shows a pronounced positive dynamics of students' academic results as a result of the introduction of digital monitoring. At the stage of input diagnostics (pre-test), the average success rate for completing tasks was 61.8 %, which reflects the basic level of mastering the content of the subject "Natural Science".

Table 1 – Dynamics of students' average academic achievement (pre-test/post-test)

Stage	Period	Average score (%)
Pre-test	January 2026	61.8
Post-test	May 2026	76.9

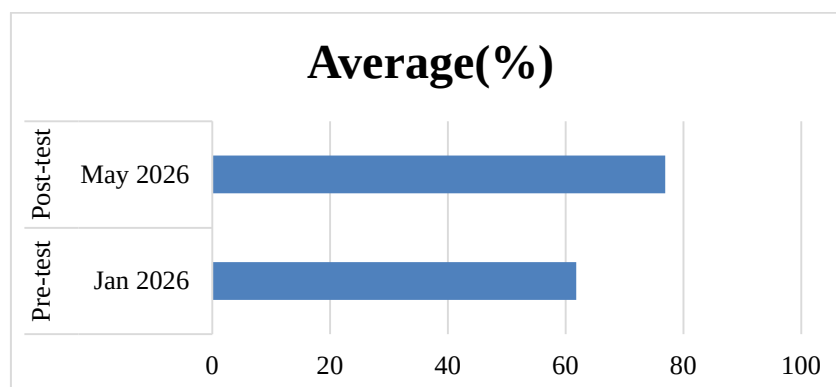


Figure 1 – Dynamics of students' average academic achievement (pre-test/post-test)

The results obtained allow us to conclude that the implemented digital monitoring model has a positive impact on students' academic performance, including in inclusive education, where individualization of the pace and trajectory of learning is especially important.

Within the framework of digital screening and monitoring of educational achievements in the Google Classroom environment, a set of diagnostic tasks on the subject of "Natural Science" was used to assess the subject knowledge, skills, and elements of educational independence of students with special educational needs. The structure of tasks included test tasks with the choice of one correct answer, tasks for establishing correspondences, short open-ended questions, as well as tasks that require analyzing visual information (diagrams, graphs) and solving simple practice-oriented learning situations. This set of tasks allowed us to assess both the reproductive level of knowledge acquisition and the basic elements of their application in academic and life contexts.

The results of the tasks were evaluated on the basis of a point-based criterion system. For test tasks, a dichotomous scale was used (1 point — correct answer, 0 points — incorrect answer). For tasks with a short and detailed answer, partial evaluation criteria were used to take into account the degree of completeness and correctness of the answer. The final result was converted to a percentage expression, which ensured comparability of data between different stages of diagnosis and the possibility of statistical processing.

The meaningful validity of evaluation tools was ensured by its development in accordance with the curriculum for natural science disciplines and the criteria for evaluating the system of summative assessment for a unit and for a term, as well as taking into account the level of cognitive goals corresponding to the Bloom taxonomy. Additionally, the tasks were evaluated by subject teachers and a specialist in special and inclusive education, after which adjustments were made to simplify the wording and reduce the mental effort for students with special educational needs.

The reliability of diagnostic instruments was ensured by preliminary pilot testing on a sample of students of a comparable level of training, which made it possible to clarify the wording of individual tasks and exclude ambiguous elements. The internal consistency of tasks was confirmed by the Cronbach coefficient, which was 0.82, denoting a high degree of reliability of the tools used. Additionally, the stability of the results was confirmed by repeated execution of individual diagnostic tasks, which showed the stability of the obtained indicators over time.

Diagnostic tools included tasks of various levels of complexity aimed at testing the assimilation of basic natural science concepts, understanding cause-and-effect relationships between natural phenomena, interpreting graphic information, and applying knowledge in practice-oriented situations. The use of the digital format provided automated registration of task results, the time of their completion, and the timeliness of sending work, which made it possible to obtain objective quantitative indicators of each student's educational activity.

At the first stage of statistical data processing, descriptive statistics were analyzed, which made it possible to assess the overall level of academic achievement of 6th grade students in the subject "Natural Science" and the degree of variability of results during digital monitoring.

The results of the input diagnostics showed that the average success rate in performing diagnostic tasks was $M = 61.8\%$ with a standard deviation of $SD = 12.4\%$, which indicates a heterogeneous level of students' training. The median value ($Me = 63.0\%$) was close to the arithmetic mean, which indicates that there is no pronounced asymmetry in the distribution of results.

After the introduction of digital monitoring on the Google Classroom system, an increase in the average level of academic achievement was noted to $M = 76.9\%$, while reducing the standard deviation to $SD = 9.1$. A decrease in the variability of results indicates a gradual leveling of students' subject preparation and a decrease in differences between students in the degree of mastering educational material. The median value also increased to $Me = 78.0\%$, which confirms the positive dynamics of the majority of study participants.

The obtained indicators allow us to conclude that the use of digital monitoring not only increased the average level of educational results, but also assisted in a more uniform development of the content of the course "Natural Science". Regular feedback, timely identification of

educational difficulties, and the possibility of prompt adjustment of educational tasks created conditions for improving the effectiveness of the educational process.

Table 2 – Descriptive statistics of the results of digital screening and monitoring of scholastic performance of 6th grade students in the subject "Natural Science" (n = 34)

Table 2 – Descriptive statistics of digital screening and monitoring results (n = 34)

Indicator	n	Min	Max	M	SD	Me
Pre-test	34	35	86	61.8	12.4	63.0
Post-test	34	55	95	76.9	9.1	78.0

Note – M = arithmetic mean; Me = median; SD = standard deviation.

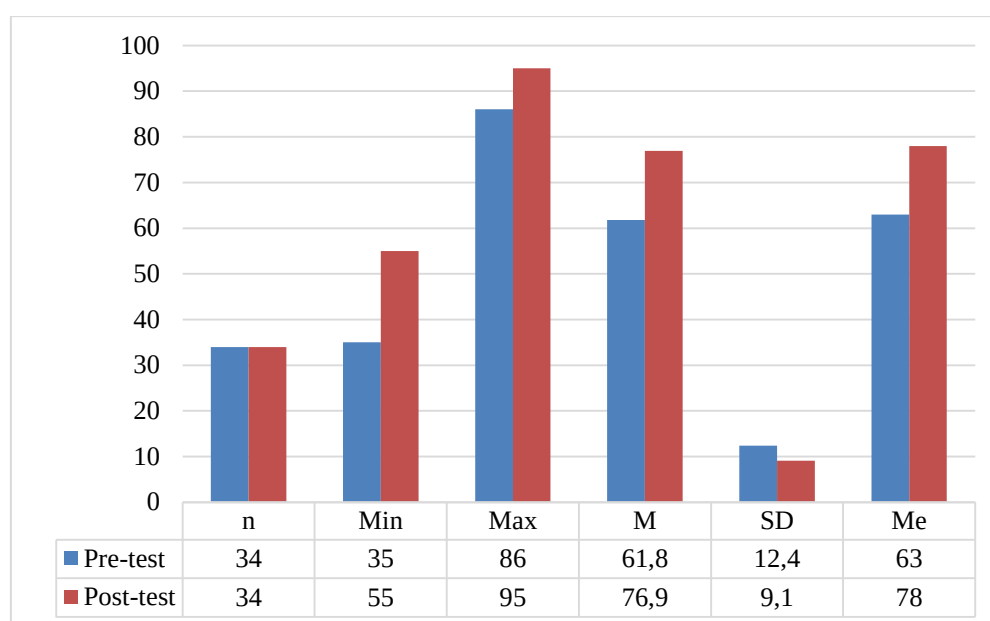


Figure 2 – Descriptive statistics of digital screening and monitoring results (n = 34)

The increase in the minimum score from 35% to 55% suggests that the most noticeable positive changes occurred among students who experienced the greatest difficulties in studying natural science. At the same time, an increase in the maximum result to 95% indicates the preservation of high educational motivation and success among students with an initially high level of training.

To evaluate the effectiveness of digital screening and learning-dynamics monitoring, a paired-samples Student's t-test was used to determine whether the pre-test and post-test results of the same group differed significantly.

The results of the analysis showed a statistically significant increase in the level of educational achievements after the introduction of digital monitoring. The average success rate of diagnostic tasks increased from $61.8 \pm 12.4\%$ to $76.9 \pm 9.1\%$.

The Student's t-test yielded $t(33) = 6.18$, with $p < 0.001$, indicating a high probability that the identified differences are due to the introduction of digital monitoring rather than random factors.

Thus, the null hypothesis about the absence of differences between the results of input and final diagnostics was rejected, and the alternative hypothesis about the positive effect of digital monitoring on learning outcomes was statistically confirmed.

The results obtained indicate that the systematic use of the Google Classroom tool for arranging digital screening, regular diagnostics, and prompt feedback contributes to improving the quality of learning material for 6th grade students. A particularly noticeable positive trend was

observed among students who experienced difficulties in studying the subject "Natural Science", which confirms the effectiveness of using digital technologies to individualize the educational process.

Table 3 - Results of checking the statistical significance of differences (paired Student's t-test)

Table 3 – Paired-samples Student's t-test results

Indicator	Before monitoring (M ± SD)	After monitoring (M ± SD)	t	df	p
Task completion success, %	61.8 ± 12.4	76.9 ± 9.1	6.18	33	< .001

The results of the study showed the presence of statistically significant relationships between the studied variables.

A strong positive correlation was found between the average score on digital diagnostic tasks and students' final academic performance ($r = 0.74$; $p < 0.001$). This indicates that students who perform better in digital tasks tend to have higher final scores in the subject of "Natural Science".

There was also a moderate negative correlation between task completion duration and academic achievement ($r = -0.52$; $p = 0.002$), which indicates that more successful students complete tasks faster and more efficiently.

In addition, a moderate positive correlation was found between the regularity of completing tasks in Google Classroom and overall scholastic performance ($r = 0.66$; $p < 0.001$), which confirms the importance of systematic learning activities for advancing educational outcomes.

Table 4-Results of Pearson correlation analysis

Table 4 – Pearson correlation analysis results

Indicator	r	p	Interpretation
Digital task scores ↔ overall performance	0.74	< .001	Strong positive relationship
Completion time ↔ performance	-0.52	.002	Moderate negative relationship
Regularity of completion ↔ performance	0.66	< .001	Moderate positive relationship

Summarizing the results of the study shows that the introduction of digital screening and systematic monitoring of the learning performance of 6th grade students in the subject "Natural Science" based on the Google Classroom framework has led to continual positive changes regarding educational results.

Descriptive statistics indicate an increase in the average level of academic achievement, a decrease in the variability of results, and a leveling of individual indicators of students, including students with special educational needs. This indicates a more uniform development of educational material and an increase in the overall effectiveness of the educational process.

The paired-samples t-test confirmed statistically significant differences between baseline and final assessment, supporting a positive association between the digital monitoring intervention and students' learning outcomes.

Pearson's correlation analysis demonstrated significant relationships between indicators of students' digital activity (task performance, regularity of learning activities, time of completion) and

their final academic performance. The most pronounced positive relationship was established between the quality of digital tasks and the level of educational achievements, which confirms the diagnostic value of students' digital traces.

Taken together, the findings show that digital screening and monitoring in Google Classroom can be used not only to assess current knowledge but also to track learning dynamics, identify difficulties, and support individualized instruction for students with SEN.

The study examined the possibilities of digital screening and monitoring of learning dynamics among 6th-grade students studying the subject "Natural Science" using the Google Classroom interface, including students with special educational needs.

The relevance of this work is determined by the rapid digitalization of education, the growing requirements for individualization of training, and the need to move from traditional final assessment to persistent monitoring of educational progress. In the context of inclusive education, early detection of learning difficulties and prompt pedagogical support are particularly important, which is impossible without using digital tools for recording and analyzing educational data. In this regard, digital screening technologies and learning analytics are becoming a key resource for improving the quality of education and guaranteeing educational equity.

The study has several limitations. It involved a relatively small sample from schools in one city and did not include a control group. Therefore, the results should be interpreted as evidence of positive change associated with the intervention rather than as definitive proof of causality. Future studies should include larger and more diverse samples, control or comparison groups, longer monitoring periods, and predictive models based on artificial intelligence.

Conclusion

The study examined the possibilities of digital screening and monitoring the dynamics of learning among 6th-grade students studying the subject "Natural Science" using the Google Classroom interface, including students with special educational needs.

The rapid digitalization of education determines the relevance of this work, the growing requirements for individualization of training, and the need to move from traditional final assessment to continuous supervision of educational progress. In the context of inclusive education, early detection of learning difficulties and prompt pedagogical support are of particular importance, which is impossible without the use of digital tools for recording and analyzing educational data. In this regard, digital screening technologies and learning analytics are becoming key resources for improving the quality of education and ensuring educational equity.

The results of the study showed that the systematic use of digital monitoring helps increase students' educational achievement and reduce the spread of individual results, indicating more uniform development of educational material. Particularly significant changes were observed in students with special educational needs, for whom digital feedback provided more stable learning dynamics along with timely correction of learning difficulties.

Statistical analysis using Student's criterion confirmed the presence of statistically significant differences between the results of the input and final diagnostics, which shows a positive impact of digital monitoring on educational results. The data obtained confirm that the transition to the systematic use of digital assessment tools can improve the objectivity of pedagogical decisions and the effectiveness of the educational process.

Pearson's correlation analysis uncovered stable relationships between indicators of students' digital activity and their academic success. The most pronounced relationship was established between the performance of digital tasks and final academic performance, which confirms the importance of digital footprints as indicators of academic success and the potential of their use in predictive analytics.

Thus, digital screening and monitoring of learning in the Google Classroom environment is an effective tool for improving the quality of science education, giving not only an assessment of the

current level of knowledge, but also support for individual educational trajectories of students, including students with special educational needs.

The practical relevance of the study lies in the possibility of introducing the developed approach into the school education system to improve the effectiveness of teaching natural science subjects, early detection of educational deficits, and strengthening targeted pedagogical support.

The prospects for further research are related to expanding the experimental base, including control groups, as well as developing predictive analytics and artificial intelligence models to predict learning dynamics and automate educational monitoring in inclusive education.

For school practice, it is recommended that teachers use a limited set of transparent digital indicators, review them regularly, combine automated data with professional pedagogical judgment, and adapt task wording, pace, and feedback to the needs of students with SEN.

References

1. Khalil M. Learning Analytics in Support of Inclusiveness and Disabled Students: A Systematic Review / M. Khalil, S. Slade, P. Prinsloo // *Journal of Computing in Higher Education*. – 2024. – Vol. 36, no. 1. – P. 129-153. – <https://doi.org/10.1007/s12528-023-09363-4>
2. Fuchs L. S. Curriculum-Based Measurement for Progress Monitoring / L. S. Fuchs, D. Fuchs // *Teaching Exceptional Children*. – 2002. – Vol. 34, no. 5. – P. 8-15.
3. Burns M. K. Reported Prevalence of Evidence-Based Instructional Practices in Special Education / M. K. Burns, J. E. Ysseldyke // *The Journal of Special Education*. – 2009. – Vol. 43, no. 1. – P. 3-11. – <https://doi.org/10.1177/0022466908315563>
4. Romero C. Educational Data Mining and Learning Analytics: An Updated Survey / C. Romero, S. Ventura // *WIREs Data Mining and Knowledge Discovery*. – 2020. – Vol. 10, no. 3. – Art. e1355. – <https://doi.org/10.1002/widm.1355>
5. Мовкебаева З. А. Развитие специального и инклюзивного образования в мире / З. А. Мовкебаева, Б. А. Дюсенбаева. – Алматы, 2021. – 168 с.
6. Абаева Г. А. Состояние и тенденции развития инклюзивного образования в Республике Казахстан / Г. А. Абаева, И. Г. Елисеева // *Педагогика и психология*. – 2023. – № 2 (55). – С. 15-24. DOI: 10.51889/2959-5762.2023.55.2.001.
7. Мовкебаева З. А. Компетенции специального педагога в инклюзивном образовании / З. А. Мовкебаева, Б. А. Дюсенбаева // *Вестник Казахского национального женского педагогического университета*. – 2021. – № 1 (85). – С. 47-55. - DOI: 10.52512/2306-5079-2021-85-1-47-55.
8. Assessing the Readiness of Future Teachers to Use Digital Technologies in Professional Activities / K. M. Berkimbayev, G. J. Niyazova, B. T. Kerimbayeva, M. M. Akeshova // *World Journal on Educational Technology: Current Issues*. – 2022. – Vol. 14, no. 5. – P. 1354-1366. - DOI: 10.18844/wjet.v14i5.7341.
9. Соловьева Т. А. Цифровая трансформация специального образования: от точечных решений к системным изменениям / Т. А. Соловьева, О. И. Кукушкина // *Альманах Института коррекционной педагогики*. – 2021. – № 44 (1). – С. 1-18.
10. Baek C. Past, Present, and Future Directions of Learning Analytics Research for Students with Disabilities / C. Baek, S. J. Aguilar // *Journal of Research on Technology in Education*. – 2023. – Vol. 55, no. 4. – P. 621–640. – DOI: <https://doi.org/10.1080/15391523.2022.2067796>
11. Mikropoulos T. A. Digital Technology Supports Science Education for Students with Disabilities: A Systematic Review / T. A. Mikropoulos, G. Iatraki // *Education and Information Technologies*. – 2023. – Vol. 28, no. 4. – P. 3911–3935. – DOI: <https://doi.org/10.1007/s10639-022-11317-9>
12. Inquiry-Based Science Instruction for Students With Disabilities: A Systematic and Meta-Analytic Review / S. E. Wilson, W. J. Therrien, J. Gersib [et al.] // *Science Education*. – 2024. – <https://doi.org/10.1002/sce.70029>
13. García Terceño E. M. Teaching Science to Students with Special Educational Needs: A Systematic Review of Science Teaching-Learning Approaches in Regular and Special Education Settings / E. M. García Terceño, I. M. Greca // *International Journal of Science Education*. – 2023. – Vol. 45, no. 12. – P. 969-989. – <https://doi.org/10.1080/09500693.2023.2179377>

14. Using Technology to Assess Individualized Education Plan Goal Progress During Virtual Learning / T. Wade, M. A. McConomy, S. Heiniger [et al.] // Journal of Special Education Technology. – 2023. – Vol. 38, no. 1. – P. 28-40. – <https://doi.org/10.1177/01626434221074054>

15. Fuchs A. Using Learning Progress Monitoring to Promote Academic Performance? A Meta-Analysis of the Effectiveness / A. Fuchs, A. Radkowsch, D. Sommerhoff // Educational Research Review. – 2025. – Vol. 46. – Art. 100648. – DOI: <https://doi.org/10.1016/j.edurev.2024.100648>

М. Хасенова, М. Смагулов, М.Еркибаева

ЖАРАТЫЛЫСТАНУ-ҒЫЛЫМИ ЦИКЛ ПӘНДЕРІН ОҚУ БАРЫСЫНДА ЕРЕКШЕ БІЛІМ БЕРУ ҚАЖЕТТІЛІКТЕРІ БАР ОҚУШЫЛАРДЫҢ ОҚУ ДИНАМИКАСЫН ЦИФРЛЫҚ СКРИНИНГТЕУ ЖӘНЕ МОНИТОРИНГТЕУ

Мақалада Қазақстандағы инклюзивті білім беру жағдайында ерекше білім беру қажеттіліктері бар (ЕББҚ) оқушылардың оқу динамикасын объективті бағалау мәселесі қарастырылады. Дәстүрлі қорытынды және критериалды бағалау көбіне соңғы нәтижені тіркейді және оқу материалын меңгерудің жеке қарқынын, когнитивтік жүктемені, қайталанатын қателерді және тапсырманы орындаудағы дербестік деңгейін толық көрсете бермейді. Бұл шектеулер абстрактілі ұғымдармен, көрнекі модельдермен, формулалармен және себеп-салдарлық байланыстармен жұмыс істеуді талап ететін жаратылыстану пәндерінде ерекше маңызды. Зерттеудің мақсаты алтыншы сыныптың инклюзивті жаратылыстану сабақтарында цифрлық скрининг пен үздіксіз педагогикалық мониторинг моделін әзірлеу және эмпирикалық тұрғыдан тексеру болды. Зерттеуге ЕББҚ бар 34 оқушы қатысты; зерттеу 2026 жылғы қаңтар-мамыр аралығында жүргізілді. Google Classroom және электрондық үлгерім журналдары арқылы тапсырма нәтижелері, орындау уақыты, әрекет саны, оқу белсенділігінің тұрақтылығы және цифрлық ресурстарды пайдалану көрсеткіштері тіркелді. Деректер сипаттамалық статистика, жұпталған Стьюдент t-критерийі және Пирсон корреляциялық талдауы арқылы өңделді. Орташа үлгерім көрсеткіші бастапқы тестілеудегі 61,8%-дан қорытынды кезеңде 76,9%-ға дейін өсті, ал нәтижелердің шашырауы азайды. Айырмашылық статистикалық тұрғыдан мәнді болды: $t(33) = 6,18$; $p < 0,001$. Цифрлық тапсырма нәтижелері мен қорытынды үлгерім арасында күшті оң байланыс анықталды ($r = 0,74$; $p < 0,001$), ал тапсырмаларды тұрақты орындау да оқу нәтижелерімен оң байланыс көрсетті ($r = 0,66$; $p < 0,001$). Нәтижелер цифрлық скрининг оқу қиындықтарын ерте анықтауға, уақтылы кері байланыс беруге және жеке білім беру траекторияларын икемді түзетуге мүмкіндік беретінін көрсетті.

Түйін сөздер: инклюзивті білім беру, ерекше білім беру қажеттіліктері, цифрлық скрининг, педагогикалық мониторинг, оқу аналитикасы, жаратылыстану-ғылыми білім беру.

М. Хасенова, М. Смагулов, М.Еркибаева

ЦИФРОВОЙ СКРИНИНГ И МОНИТОРИНГ ДИНАМИКИ ОБУЧЕНИЯ УЧАЩИХСЯ С ОСОБЫМИ ОБРАЗОВАТЕЛЬНЫМИ ПОТРЕБНОСТЯМИ ПРИ ИЗУЧЕНИИ ПРЕДМЕТОВ ЕСТЕСТВЕННОНАУЧНОГО ЦИКЛА

В статье рассматривается проблема объективной оценки динамики обучения учащихся с особыми образовательными потребностями (ООП) в условиях инклюзивного образования Казахстана. Традиционное итоговое и критериальное оценивание преимущественно фиксирует конечный результат и не всегда отражает индивидуальный темп усвоения материала, когнитивную нагрузку, повторяющиеся ошибки и степень самостоятельности учащегося. Эти ограничения особенно значимы при изучении естественнонаучных дисциплин, предполагающих работу с абстрактными понятиями, визуальными моделями, формулами и причинно-следственными связями. Цель исследования заключалась в разработке и эмпирической проверке модели цифрового скрининга и непрерывного педагогического мониторинга в инклюзивных классах шестого класса. В исследовании приняли участие 34 учащихся с ООП; работа проводилась с января по май 2026 года. С использованием Google Classroom и электронных журналов фиксировались результаты заданий,

время выполнения, количество попыток, регулярность учебной активности и обращения к цифровым ресурсам. Данные анализировались методами описательной статистики, парного t-критерия Стьюдента и корреляционного анализа Пирсона. Средний показатель успеваемости увеличился с 61,8% на входном тестировании до 76,9% на итоговом этапе, а вариативность результатов снизилась. Различия были статистически значимыми: $t(33) = 6,18$; $p < 0,001$. Между результатами цифровых заданий и итоговой успеваемостью выявлена сильная положительная связь ($r = 0,74$; $p < 0,001$), а регулярность выполнения заданий также положительно коррелировала с результатами обучения ($r = 0,66$; $p < 0,001$). Полученные данные показывают, что цифровой скрининг способствует раннему выявлению трудностей, своевременной обратной связи и гибкой корректировке индивидуальных образовательных траекторий.

Ключевые слова: инклюзивное образование, особые образовательные потребности, цифровой скрининг, педагогический мониторинг, учебная аналитика, естественнонаучное образование.

References

1. Khalil, M., Slade, S., & Prinsloo, P. (2024). Learning analytics in support of inclusiveness and disabled students: A systematic review. *Journal of Computing in Higher Education*, 36(1), 129-153. <https://doi.org/10.1007/s12528-023-09363-4>
2. Fuchs, L. S., & Fuchs, D. (2002). Curriculum-based measurement for progress monitoring. *Teaching Exceptional Children*, 34(5), 8-15.
3. Burns, M. K., & Ysseldyke, J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *The Journal of Special Education*, 43(1), 3-11. <https://doi.org/10.1177/0022466908315563>
4. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), Article e1355. <https://doi.org/10.1002/widm.1355>
5. Movkebayeva, Z. A., & Dyusenbayeva, B. A. (2021). *Razvitiye spetsial'nogo i inklyuzivnogo obrazovaniya v mire* [Development of special and inclusive education in the world]. Almaty. 168 p.
6. Abayeva, G. A., & Yelisseyeva, I. G. (2023). Sostoyaniye i tendentsii razvitiya inklyuzivnogo obrazovaniya v Respublike Kazakhstan [State and development trends of inclusive education in the Republic of Kazakhstan]. *Pedagogika i psikhologiya [Pedagogy and Psychology]*, 2(55), 15-24. DOI: 10.51889/2959-5762.2023.55.2.001.
7. Movkebayeva, Z. A., & Dyusenbayeva, B. A. (2021). Kompetentsii spetsial'nogo pedagoga v inklyuzivnom obrazovanii [Competencies of a special educator in inclusive education]. *Vestnik Kazakhskogo natsional'nogo zhenskogo pedagogicheskogo universiteta [Bulletin of the Kazakh National Women's Teacher Training University]*, 1(85), 47-55. DOI: 10.52512/2306-5079-2021-85-1-47-55.
8. Berkimbayev, K. M., Niyazova, G. J., Kerimbayeva, B. T., & Akeshova, M. M. (2022). Assessing the readiness of future teachers to use digital technologies in professional activities. *World Journal on Educational Technology: Current Issues*, 14(5), 1354-1366. DOI: 10.18844/wjet.v14i5.7341.
9. Solov'yeva, T. A., & Kukushkina, O. I. (2021). Tsifrovaya transformatsiya spetsial'nogo obrazovaniya: Ot tochechnykh resheniy k sistemnym izmeneniyam [Digital transformation of special education: From point solutions to systemic changes]. *Al'manakh Instituta korrektsionnoy pedagogiki [Almanac of the Institute of Special Education]*, 44(1), 1-18.
10. Baek, C., & Aguilar, S. J. (2023). Past, present, and future directions of learning analytics research for students with disabilities. *Journal of Research on Technology in Education*, 55(4), 621-640. <https://doi.org/10.1080/15391523.2022.2067796>
11. Mikropoulos, T. A., & Iatraki, G. (2023). Digital technology supports science education for students with disabilities: A systematic review. *Education and Information Technologies*, 28(4), 3911-3935. <https://doi.org/10.1007/s10639-022-11317-9>
12. Wilson, S. E., Therrien, W. J., Gersib, J., et al. (2024). Inquiry-based science instruction for students with disabilities: A systematic and meta-analytic review. *Science Education*. <https://doi.org/10.1002/sce.70029>
13. García Terceño, E. M., & Greca, I. M. (2023). Teaching science to students with special educational needs: A systematic review of science teaching-learning approaches in regular and special

education settings. *International Journal of Science Education*, 45(12), 969-989. <https://doi.org/10.1080/09500693.2023.2179377>

14. Wade, T., McConomy, M. A., Heiniger, S., et al. (2023). Using technology to assess Individualized Education Plan goal progress during virtual learning. *Journal of Special Education Technology*, 38(1), 28-40. <https://doi.org/10.1177/01626434221074054>

15. Fuchs, A., Radkowsch, A., & Sommerhoff, D. (2025). Using learning progress monitoring to promote academic performance? A meta-analysis of the effectiveness. *Educational Research Review*, 46, Article 100648. <https://doi.org/10.1016/j.edurev.2024.100648>

Авторлар туралы мәлімет:

Хасенова Марал - PhD, аға оқытушы, Ө. Марғұлан атындағы Павлодар педагогикалық университеті, Павлодар, Қазақстан, e-mail: maral_176@mail.ru

Смагулов Мейрман (автор-корреспондент) - докторант, Ө. Марғұлан атындағы Павлодар педагогикалық университеті, Павлодар, Қазақстан, e-mail: meirmansmagulov@gmail.com

Еркибаева Меруерт - химия ғылымдарының кандидаты, қауымдастырылған профессор, Ө. Марғұлан атындағы Павлодар педагогикалық университеті, Павлодар, Қазақстан, e-mail: erkibaevameruert@mail.ru

Сведения об авторах:

Хасенова Марал - PhD, старший преподаватель, Павлодарский педагогический университет имени Ө. Марғұлан, Павлодар, Казахстан, e-mail: maral_176@mail.ru

Смагулов Мейрман (автор-корреспондент) - докторант, Павлодарский педагогический университет имени Ө. Марғұлан, Павлодар, Казахстан, e-mail: meirmansmagulov@gmail.com

Еркибаева Меруерт - кандидат химических наук, ассоциированный профессор, Павлодарский педагогический университет имени Ө. Марғұлан, Павлодар, Казахстан, e-mail: erkibaevameruert@mail.ru

Information about authors:

Khassenova Maral - PhD, Senior Lecturer, A. Margulan Pavlodar Pedagogical University, Pavlodar, Kazakhstan, e-mail: maral_176@mail.ru

Smagulov Meirman (corresponding author) - PhD Student, A. Margulan Pavlodar Pedagogical University, Pavlodar, Kazakhstan, e-mail: meirmansmagulov@gmail.com

Erkibayeva Meruert - Candidate of Chemical Sciences, Associate Professor, A. Margulan Pavlodar Pedagogical University, Pavlodar, Kazakhstan, e-mail: erkibaevameruert@mail.ru

А.Б. Ертілеуова^{1*}, Р. Желвис²¹Ш. Есенов атындағы Каспий технологиялар және инжиниринг университеті, Актау қ., Қазақстан Республикасы²Вильнюс университеті, Философия факультеті, Білім беру ғылымдары институты, Вильнюс қ., Литва

*e-mail: vertileuovaaida@gmail.com

¹ORCID 0009-0004-3392-1719, ²ORCID 0000-0002-2757-5651

БОЛАШАҚ ПЕДАГОГТАРДЫҢ КӘСІБИ ДАЯРЛЫҒЫНДА МЕКТЕП ТОП-МЕНЕДЖЕРІНІҢ САУАТТЫЛЫҒЫН ҚАЛЫПТАСТЫРУ

Зерттеуде болашақ педагогтарды кәсіби даярлау үдерісінде мектеп топ-менеджерінің сауаттылығын қалыптастыру мәселесі қарастырылады. Қазіргі білім беру жүйесіндегі басқару қызметінің күрделенуі, білім беру ұйымдарының бәсекеге қабілеттілігін арттыру қажеттілігі және педагог кадрларға қойылатын жаңа кәсіби талаптар болашақ мұғалімдердің басқарушылық құзыреттерін дамытудың өзектілігін арттырып отыр. Зерттеудің мақсаты - болашақ педагогтардың кәсіби даярлығы барысында мектеп топ-менеджерінің сауаттылығын қалыптастырудың теориялық және әдіснамалық негіздерін айқындау, оның құрылымдық компоненттерін жүйелеу және жоғары білім беру жағдайында аталған сапаны дамытудың педагогикалық мүмкіндіктерін негіздеу. Зерттеу аясында мектеп топ-менеджерінің сауаттылығы білім беру ұйымын тиімді басқаруға қажетті стратегиялық ойлау, көшбасшылық, басқарушылық шешім қабылдау, коммуникациялық мәдениет, командалық өзара әрекет, өзгерістерді басқару және ұйымдастырушылық қызметті жоспарлау құзыреттерінің өзара байланысқан жүйесі ретінде қарастырылады. Отандық оқу бағдарламалары мұғалімді пәндік-әдістемелік тар шеңберде қалдырса, шетелдік тәжірибе оған бүкіл мекеменің стратегиялық шешімдеріне тікелей ықпал ететін жауапты менеджер рөлін жүктейді. Сол себепті білім беру менеджменті мен көшбасшылық саласындағы тұжырымдамалар салыстырылып, мектеп топ-менеджерінің функционалдық құрылымы педагогтің кәсіби даярлық деңгейімен нақты байланыстырылды. Басты мәселе - осы. Басқарушылық сауаттылық кабинеттік нұсқаулықтармен ғана өлшенбейді. Сараптама көрсеткендей, жүйенің тірегі стратегиялық пайым, шешім қабылдау, командалық үйлесім және ұйымдастырушылық сияқты алты нақты құзыретке саяды. Тек сабақ берушіні даярлау басқару вакуумын толтыра алмайды. Сондықтан университет бағдарламаларына білім беру менеджменті модульдерін қосып, нақты мектептік жанжалдар мен ресурстық тәуекелдерге негізделген практикалық кейстерді көбейту қажеттілігі негізделді. Маман шешім қабылдауы керек. Педагогты ұйымды басқару үдерістеріне қатыса алатын деңгейге жеткізбейінше сапалық ілгерілеу болмайды, ал ұсынылған тұжырымдар ЖОО бағдарламаларын жаңартуға ғылыми-әдістемелік тірек береді.

Түйін сөздер: болашақ педагог, кәсіби даярлық, мектеп топ-менеджері, басқарушылық сауаттылық, басқарушылық құзырет, көшбасшылық, білім беру менеджменті.

Кіріспе

Мектеп қабырғасына енді қадам басқан жас мұғалім көбіне 45 минуттық сабақтан асып, ұйымдастырушылық дағдарыстар мен ата-аналар наразылығын шешуге қауқарсыз болып шығады. Мәселе осында. Қазақстанның орта білім беру ұйымдарында адами капиталды дамыту мен басқару тиімділігіне мемлекеттік деңгейде басымдық берілгенімен, университет аудиторияларында мұғалім әлі де тек пәндік орындаушы ретінде даярланып келеді. Бұл жүйе ескірді. Заманауи білім ошағына сабақ өткізуші емес, командалық жұмысты үйлестіріп, өзгерістерді тізгіндей алатын ұйымдастырушы маман ауадай қажет. Басқарушылық сауаттылық педагогикалық ЖОО-ларда базалық деңгейде берілуі шарт, әйтпесе мектептердегі кадрлық сапа дағдарысы тереңдей береді.

Цифрлық есеп беру мен академиялық еркіндік қатар енгізілгенімен, жас мамандарда стратегиялық жоспарлау дағдысы тапшы. Талап күшейді. Алайда педагогикалық

бағдарламаларда көшбасшылық пен менеджментке арналған ғылыми-әдістемелік негіздердің жеткіліксіздігі реформалардың төменгі буында тұралап қалуына әкеліп соғады.

Зерттеу мақсаты - болашақ педагогтардың кәсіби даярлығы аясында мектеп топ-менеджерінің сауаттылығын қалыптастырудың теориялық негіздерін дәйектеу, оның құрылымдық компоненттерін анықтау және тиімді педагогикалық шарттарын нақтылау.

Зерттеу міндеттері:

- мектеп топ-менеджерінің сауаттылығы ұғымының теориялық негіздерін сараптау;
- басқарушылық сауаттылықты қалыптастырудың педагогикалық шарттары мен нақты әдістерін дәйектеу;
- кәсіби даярлау үдерісінде қолданылатын басқарушылық сауаттылық моделін әзірлеу.

Жоғары оқу орындарында стандартты бағдарламаларды қайталау нәтиже бермейді. Егер педагог даярлау жүйесіне мектеп топ-менеджерінің сауаттылығына бағытталған арнайы модель енгізіліп, білім беру мазмұны инновациялық технологиялармен, нақты мектептік жанжалдарды шешуге арналған жобалық тапсырмалармен және рефлексивтік бағалаумен кешенді түрде күшейтілсе, онда болашақ мамандардың көшбасшылық, стратегиялық пайым және шешім қабылдау сапасы дағдарыстық жағдайларда айтарлықтай артады деген ғылыми болжам жасалады. Тәуекел көп, бірақ бұл тәсілсіз басқарушылық сауаттылықты нақты өмірмен ұштастыру мүмкін емес.

Зерттеу материалдары мен әдістері

Білім беру ұйымдарын тек сабақ кестесін орындайтын мекеме деп санау өзгерді. Бүгінгі мектепке инновация енгізіп, дағдарыстық өзгерістерді тізгіндей алатын адами капитал орталығы қажет. Бірақ директордың жеке дара бұйрығына сүйенген әкімшілік жүйе заманауи ұйымдастырушылық сын-қатерлерге төтеп бере алмай жатыр. Басқарудың командалық моделіне көшу қажеттілігі осыдан шықты. Мұғалімді пәндік-әдістемелік тар шеңберде ұстау мүмкін емес, оның басқарушылық әлеуетін ашу басты талапқа айналды. Мектеп топ-менеджерінің сауаттылығын ЖОО қабырғасында қалыптастыру бүгінгі педагогикалық ғылым мен тәжірибенің шешілмеген түйіні [1]. Бұл жерде топ-менеджмент кабинеттік ережелер жиынтығы емес. Ол - ұжымдық ресурсты жұмылдырып, ортақ шешім қабылдауға бағытталған нақты басқару технологиясы.

Мектепті жалғыз басшы сүйрей алмайды. Бұл - нақты факт. Ұйымның тұрақты дамуы директордың жеке шешімдеріне емес, ұжымдық жауапкершілік пен командалық әрекетке тікелей тәуелді [2]. Топ-менеджер рөлі жоғары лауазыммен шектелмейді. Ол - ұжымды ортақ мақсатқа бағыттап, басқару үдерісіне белсенді қатысатын кәсіби маман. Тәжірибеде бұл қызметті директордың орынбасары, әдістемелік бірлестік төрағасы, шығармашылық топ жетекшісі, кәсіби қоғамдастық үйлестірушісі немесе тәжірибелі мұғалім де еркін атқара алады.

Классикалық менеджменттегі жоспарлау, үйлестіру, ынталандыру, бақылау және бағалау сияқты қатаң функциялар қазіргі мектеп ортасында бастапқы түрінде жұмыс істемейді [3]. Білім ошағына ұжымдық көшбасшылық, серіктестік, кәсіби сенім және ортақ жауапкершілік қағидаттары көбірек қажет. Аталған тетіктер іс жүзінде сақталмаса ше? Мектептің 5 жылдық даму жоспары сөредегі шаң басқан есеп болып қалу тәуекеліне ұшырайды. Сенім мен серіктестік орнаған жерде ғана педагогтер кәсіби өсіп, білім сапасы тұрақты сақталады.

«Педагог» кәсіби стандарты мен білім саласындағы нормативтік құжаттар мұғалімнен тек оқытуды ғана емес, ұйымдастырушылық пен көшбасшылықты қатар талап етеді [3]. Талап қатаң. Алайда педагогикалық жоғары оқу орындарының білім беру бағдарламалары бұл қарқынға әлі ілесе алмай отыр. Бағдарламалар кешегі күннің өлшемімен жасалған. Оқу жоспарларында командалық басқаруға арналған қолданбалы модульдердің қарастырылмауы

салдарынан жас маман мектепке келген алғашқы айда-ақ ұйымдастырушылық вакуумға тап болады.

Ғылыми әдебиеттерді сүзгіден өткізгенде басқару теориясы, стратегиялық жоспарлау, көшбасшылық, тиімді коммуникация, командалық мәдениет, педагогикалық шешім қабылдау және кәсіби рефлексия негізгі құрылымдық компоненттер ретінде сұрыпталды. Осы элементтердің бірлігі болашақ маманға мектептің стратегиялық мақсаттарын жүзеге асыруға жол ашады. Бірақ теорияда дәйектелген құзыреттерді аудиторияда қалай қалыптастырып, қандай критерийлермен өлшеу керек деген мәселе әлі де пікірталас тудырып отыр. Ортақ байлам жоқ. Педагогтардың басқарушылық даярлығын тәжірибеге бейімдейтін нақты модельдердің ғылыми негіздемесі әлі де жетілдіруді қажет етеді [4]. Дәл осы түйткіл зерттеудің басты ғылыми жаңалығы мен практикалық маңыздылығын айқындайды.

Мұғалімнің ішкі әлеуетін ашпайынша, мектептің стратегиялық даму жоспары тек қағаз жүзіндегі есеп болып қалады. Тәжірибеде топ-менеджмент кабинеттік нұсқаулықтарға емес, нақты жұмыс істейтін жеті қағидатқа сүйенеді: ортақ мақсаттың болуы, командалық жауапкершілік, ашық коммуникация, педагогтердің кәсіби ынтымақтастығы, шешімдерді бірлесе қабылдау, үздіксіз кәсіби даму және өзгерістерді басқару. Мұндай тәсіл мектеп ішіндегі әкімшілік қысымды азайтып, білім сапасына тікелей ықпал етеді [5].

Іс жүзінде топ-менеджер бес негізгі функцияны қатар атқарады:

Жоспарлау. Педагогикалық ұжымның нақты мақсаттарын бекіту, іс-шаралар кезеңін белгілеу және күтілетін нәтижені алдын ала болжау.

Ұйымдастыру. Мұғалімдердің жүктемесін үйлестіру, міндеттерді нақты бөлу, жұмыс топтарын жасақтау және бірлескен жобаларды басқару.

Көшбасшылық. Ұжымда сенімге негізделген қарым-қатынас орнату, әріптестерге кәсіби қолдау көрсету және дағдарыстық өзгерістерді дер кезінде реттеу.

Бақылау және бағалау. Оқу-тәрбие нәтижелерін құрғақ цифрмен емес, мазмұндық тұрғыдан талдау, кері байланыс тетіктерін ұйымдастыру және атқарылған жұмыстың нақты тиімділігін сараптау.

Коммуникация. Әкімшілік, мұғалімдер, ата-аналар комитеті мен салалық органдар арасындағы ақпараттық алшақтықты жою.

Университет қабырғасында басқару дағдысын меңгермеген жас маман мектепке келген алғашқы айларда-ақ ұжымдық жанжалдар алдында қауқарсыз қалады. Бұл анық тәуекел. Осы олқылықты жою үшін болашақ мұғалімнің даярлығына басқарушылық құзыреттерді жүйелі енгізу қажет. Білім беру менеджментінің классикалық негіздерін қалаған Peter F. Drucker, Henry Mintzberg және Michael Fullan басқару тетіктері көшбасшылық пен өзгерістерді басқару қағидаттарына тікелей тәуелді екенін бұрыннан дәйектеген [6].

Бұл ұстаным мектеп қабырғасында қалай көрініс табады? Alma Harris, Kenneth Leithwood және Viviane Robinson өз зерттеулерінде білім сапасы мұғалімнің көшбасшылық әлеуетіне тікелей байланғанын көрсетеді. Демек, педагогты сабақ кестесін орындаушы ретінде ғана емес, мектептегі реформалардың тікелей бастамашысы ретінде қарастыру қажет.

Теориялық білім нақты ұйымдастырушылық әрекетпен бекітілмесе, нәтиже бермейді. Linda Darling-Hammond, Lee Shulman және John Hattie мұғалімнің кәсіби шеберлігі басқарушылық, коммуникативтік және рефлексивтік дағдылардың тоғысуынан құралатынын дәлелдейді [7]. Отандық ғылыми орта да бұл мәселеден тыс қалған жоқ. Қазақстандық зерттеушілер білім беру ұйымдарындағы басқару мәдениетін ұлттық ерекшеліктерге сай бейімдеу бағытында маңызды тұжырымдар жасады. Т.М. Баймолдаев педагогикалық менеджментті барлық қатысушылардың өзара ынтымақтастығына сүйенетін демократиялық модель ретінде түсіндіреді [8]. Ғалымның бұл пайымы болашақ мамандардың басқарушылық негізін қалыптастыруға нақты ғылыми тірек береді.

С.М. Бахишева заманауи мектептегі инновациялық менеджменттің маңызын алдыңғы қатарға шығарады. Ұйымды тиімді басқару стратегиялық жоспарлауға, ұжымның кәсіби

жауапкершілігіне және ортақ көшбасшылыққа тікелей тәуелді [9]. Бұл түс іс жүзінде әлі де күмән тудырады, себебі көптеген мектептерде инновация жарияланғанымен, ескі әкімшілік әдістер әлі сақталып отыр.

Н.С. Әлқожаева мен Қ.Б. Жұмабекова басқаруды мектептің барлық субъектілерінің өзара әрекеттесуін реттейтін біртұтас жүйе деп бағалайды [10]. Авторлар ЖОО аудиториясында басқарушылық пен ұйымдастырушылық қабілеттерді кешенді түрде ұштастыруды талап етеді. Дәл осы ойды Ш.Х. Құрманалина да қуаттайды: мұғалімнің құзыреттілігі тек пәндік біліммен шектелмейді, оған коммуникативтік және ұйымдастырушылық біліктер тікелей кіреді [11]. Осы ұстанымдар болашақ педагогтың бойында мектеп топ-менеджеріне тән сауаттылықты қалыптастырудың негізгі іргетасы болып саналады.

Ғылыми әдебиеттерді сүзгенде мектеп топ-менеджерінің сауаттылығы мотивациялық, когнитивтік, іс-әрекеттік және рефлексивтік төрт тірекке сүйенетіні көрінді. Алайда теорияда аталған осы компоненттерді нақты біріктіретін қолданбалы модельдер жоқтың қасы [12]. Жүйесіздік айқын. Көптеген зерттеулер басқарушылық даярлықты ұсынумен шектеледі, бірақ оны аудиторияда қалай құрылымдау керектігін нақтыламайды.

Дәстүрлі дәрістермен көшбасшы тәрбиелеу - уақыт жоғалту. Егер оқу үдерісіне кейс-әдіс, жобалық технологиялар, басқарушылық симуляциялар мен рефлексивтік талдау кешенді енгізілсе, болашақ мамандардың төрт компонент бойынша көрсеткіштері сапалық өзгеріске ұшырайды деген ғылыми болжам қалыптасты [13]. Бұл тәсіл әдіснаманың өзегіне айналды. Мұғалім тек сабақ өткізуші емес, мектептегі дағдарыстық үдерістерді тізгіндейтін менеджер ретінде қалыптасуы шарт.

Педагогикалық модельдің өміршеңдігін құр пайыммен дәлелдеу қиын. Сондықтан ұсынылған технологиялардың нақты нәтижесін тексеру үшін эксперименттік және бақылау топтары жасақталып, олардың бастапқы және соңғы даярлық деңгейлері бетпе-бет салыстырылды.

Талдау нәтижесі басқарушылық сауаттылықты өлшейтін төрт көрсеткіштің нақты өзгерісін ашты. Алынған деректер теориялық қағидаларды нақтылап қана қоймай, педагогикалық ЖОО бағдарламаларын жаңартуға қолданбалы негіз ұсынады. Тұжырымдар нақты.

Зерттеу аралас әдіснамаға (mixed methods) негізделді. Бұл таңдау себепсіз емес: тек сандық көрсеткіштер студенттің ішкі көшбасшылық динамикасын толық көрсетпейді, ал таза сапалық сипаттамада дәлдік жетіспейді [15]. Қос тәсілдің ұштасуы басқарушылық сауаттылыққа ықпал ететін жасырын факторларды ашуға мүмкіндік берді.

Зерттеу квазиэксперименттік дизайн арқылы ұйымдастырылды. Барысы: алдын ала диагностика (pre-test), қалыптастыру кезеңі және қорытынды өлшеу (post-test). Модельдің қаншалықты тиімді екені осы кезеңдік салыстырудан кейін ғана айқындалды.

Бастапқы деңгейді анықтау үшін онлайн форматтағы арнайы сауалнама таңдалды. Интернет құралдары қатысушылардың шынайы пікірін анонимді сақтауға және деректерді жедел жинақтауға ыңғайлы.

Сынаққа 2-4 курс аралығында оқитын 120 студент ерікті түрде тартылды. Қатысушылар алдын ала этикалық қағидалармен және мәліметтердің құпиялылығымен толық таныстырылды. Репрезентативтілік қатаң сақталды.

Топтар теңдей бөлінді: 60 студент эксперименттік топқа, қалған 60 студент бақылау тобына енді. Бақылау тобы дәстүрлі бағдарламамен жүрсе, эксперименттік аудиторияға мектептегі нақты дағдарыстар мен басқарушылық тапсырмаларға негізделген тәжірибелік шаралар берілді. Айырмашылық осы тәсілде.

Эмпирикалық кезеңдегі сұрақнама Google Forms платформасы арқылы жүргізілді. Сауалнама сілтемесі 120 қатысушыға корпоративтік пошталар, студенттік ресми мессенджерлер мен университет платформалары арқылы таратылып, қайтарым көрсеткіші дер кезінде тіркелді.

Диагностикалық құрал 24 сұраққа сығымдалып, мектеп топ-менеджерінің сауаттылығын айқындайтын төрт негізгі бағытқа теңдей бөлінді. Әр блокқа 6 сұрақтан бекітілді: мотивациялық, когнитивтік, іс-әрекеттік және рефлексивтік өлшемдер. Мұндай пропорция балансты сақтау үшін таңдалды. Бағалау классикалық 5 балдық Лайкерт шкаласы арқылы жүргізілді: «толық келісемін» - 5 балл, «келісемін» - 4 балл, «нақты пікірім жоқ» - 3 балл, «келіспеймін» - 2 балл, «мүлдем келіспеймін» - 1 балл. Студенттердің жеке ұпайлары осы шкала бойынша жинақталып, жеке нәтижесі есептелді. Респондент жинай алатын ең жоғары меже - 120 балл, ал ең төменгі шек 24 балды көрсетті.

Мәліметтерді өңдеу қолмен атқарылған жоқ. Сауалнама қорытындысы Google Forms сервисінен Microsoft Excel ортасына түсіріліп, статистикалық терең өңдеу IBM SPSS Statistics 27.0 пакеті арқылы жүзеге асырылды. Пайыздық үлес жауап бергендердің нақты санын (n) жалпы іріктеме көлеміне ($N = 120$) бөліп, жүзге көбейту арқылы есептелді: $P = (n / N) \times 100\%$.

Есептеу логикасы айқын. Мәселен, белгілі бір дағды бойынша 64 студент жоғары деңгей көрсетсе, бұл жалпы топтың 53,3%-ын құрайды. Орташа деңгейде қалған 48 респонденттің үлесі тура 40,0%-ды берсе, төмен нәтиже көрсеткен 8 студент бар болғаны 6,7%-ға тең болды.

Дәл осы математикалық алгоритм төрт көрсеткіштің әрқайсысына жеке қолданылып, эксперименттік және бақылау топтарының алшақтығы салыстырылды. Нәтижелер графикалық форматта визуализацияланып, статистикалық тексеруден өтті. Сауалнама мазмұны көшбасшылық, командалық жұмыс, басқарушылық шешім қабылдау және мектептегі дағдарыстық тәуекелдерді еңсеру дайындығын жан-жақты бағалауға мүмкіндік берді [17]. Теория осылайша нақты цифрға ұласты.

Эксперимент бірден басталған жоқ, ол өзара сабақтас үш нақты кезеңге бөлінді. Бастапқы кезеңде отандық және шетелдік авторлардың педагогикалық менеджмент пен көшбасшылыққа қатысты еңбектері, сондай-ақ Қазақстан Республикасының қолданыстағы нормативтік-құқықтық базасы мен «Педагог» кәсіби стандарты терең сүзгіден өтті [18]. Заңнама мен тәжірибе арасындағы алшақтық осы тұста нақтыланды. Бұл талдау мектеп топ-менеджерінің сауаттылығы ұғымын абстрактілі түсініктен арылтып, оның құрылымдық компоненттерін нақты сипаттауға мүмкіндік берді.

Екінші кезеңде студенттердің бастапқы даярлық деңгейін өлшеу үшін кешенді диагностикалық құралдар кешені іске қосылды [19]. Жай ғана сауалнамамен шектелу шынайы көріністі бермейтіні белгілі. Сондықтан авторлық сауалнамамен қатар өзін-өзі бағалау парақтары, сараптамалық карталар, бақылау хаттамалары және жартылай құрылымданған сұхбат әдістері қатар қолданылды.

Авторлық сауалнама студенттердің стратегиялық жоспарлау, ұжымдық шешім қабылдау, коммуникация және басқарушылық жауапкершілік деңгейін тікелей тексеруге бағытталды [20]. Мұндағы сұрақтардың барлығы бес балдық Лайкерт шкаласына негізделді. Бағалаудың объективтілігін күшейту үшін аудиториялық топтық жұмыстарға сырттай педагогикалық бақылау жасалып, оған білім беру менеджменті саласында тәжірибесі бар білікті оқытушылар сарапшы ретінде тартылды.

Үшінші кезең - негізгі сынақ алаңы. Эксперименттік топқа құзыреттілік пен тәжірибеге бағдарланған қағидаттарды біріктіретін мектеп топ-менеджерінің арнайы моделі енгізілді.

Қалыптастыру кезеңінде құрғақ дәрістер толықтай қолданбалы әдістермен алмастырылды:

- жобалық оқыту технологиясы;
- кейс-стади;
- басқарушылық рөлдік және іскерлік ойындар;
- мектептік дағдарыстық жағдаяттарды шешу;
- стратегиялық SWOT және PEST талдау элементтері;
- командалық жобаларды әзірлеу;
- цифрлық білім беру платформаларын пайдалану;

- рефлексивтік күнделік жүргізу.

Студенттерге мектептің нақты даму жоспарын түзу, педагогикалық кеңестің күн тәртібін бекіту, ата-аналармен даулы мәселелерді реттеу және қаржылық-ұйымдастырушылық тәуекелдерді есептеу тапсырылды. Бұл жай ғана жаттығу емес. Нақты мектептік шиеленістерді шешу арқылы болашақ мамандардың басқарушылық дағдылары шыңдалды.

Бақылау тобында ешқандай жаңа әдістемелік эксперимент жүргізілген жоқ; олар бекітілген типтік бағдарламамен дәстүрлі түрде оқытылды. Айырмашылық осы екі ортаның қарама-қайшылығында жатыр.

Қорытынды өлшеу кезінде басқарушылық сауаттылық төрт негізгі өлшемге бөлінді: мотивациялық, когнитивтік, іс-әрекеттік және рефлексивтік. Әр компонент жоғары, орта және төмен деңгейлерге сұрыпталды.

Эмпирикалық мәліметтер IBM SPSS Statistics 27.0 пакетінде өңделіп, нәтижелер Стьюденттің тәуелсіз t-критерийі мен Пирсонның χ^2 критерийі арқылы салыстырылды [21]. Сауалнаманың ішкі үйлесімділігі Cronbach's Alpha коэффициенті бойынша тексеріліп, оның көрсеткіші $\alpha = 0,87$ мәнін берді. Бұл зерттеу құралының статистикалық тұрғыдан өте сенімді екенін растайды. Барлық қатысушылар процеске ерікті түрде қосылып, ақпараттандырылған келісім ұсынды, ал алынған дербес деректер қатаң құпия сақталып, тек ғылыми қорытындылар шығару үшін пайдаланылды.

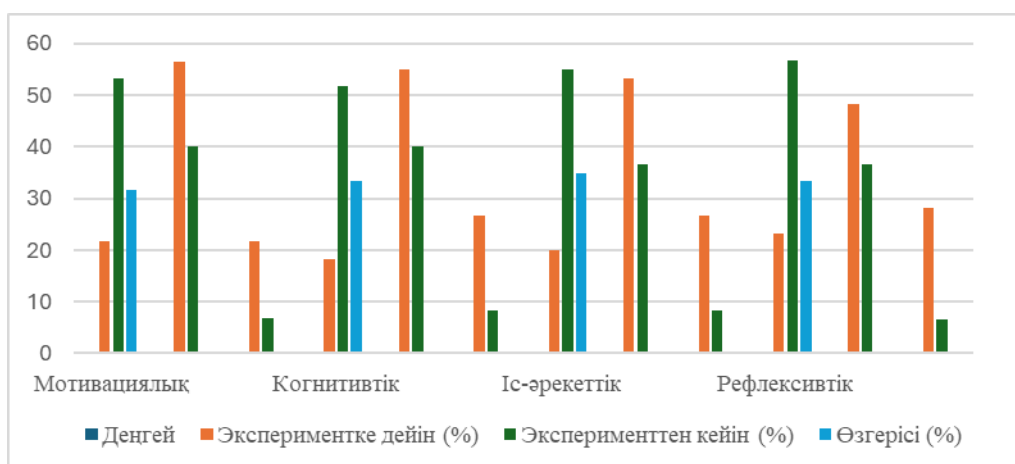
Нәтижелер және талқылау

Бастапқы диагностика қос топтың мүмкіндігі бірдей екенін көрсетті. Тәжірибеге дейін екі топтың нәтижелері арасында айтарлықтай статистикалық айырмашылық байқалған жоқ ($p > 0,05$). Алайда нақты мектептік жанжалдарды реттеу мен басқарушылық шешім қабылдау симуляциялары оқу процесіне енгізілген соң көрсеткіштер сапалық өзгеріске түсті. Серпіліс бірден сезілді. Мотивациялық блок бойынша өзін мектеп басшылығы немесе пәндік бірлестік жетекшісі ретінде көре алатын студенттер үлесі 21,7%-дан бірден 53,3%-ға дейін күрт көтерілді. Жауапкершіліктен қашатындар азайды. Бұрын көшбасшылыққа күмәнмен қараған орта буын 56,6%-дан 40,0%-ға ығысса, бастама көтеруден тартынатын төменгі топ 21,7%-дан 6,7%-ға дейін қысқарды. Бұл тек баға үшін емес. Мамандардың мектептегі басқарушылық үдерістерді тізгіндеуге деген шынайы қызығушылығы оянды.

Когнитивтік өлшем бойынша да айқын өсім тіркелді. Білім беру ұйымының нормативтік құжаттары мен қаржылық механизмдерін терең меңгерген жоғары деңгейдегі респонденттер үлесі 18,3%-дан 51,7%-ға дейін артты. Мәселе құжатта емес. Бұрын мектеп менеджментін үстірт түсініп келген орта топ 55,0%-дан 40,0%-ға түссе, базалық білімі жеткіліксіз әлсіз топтағылар 26,7%-дан 8,3%-ға дейін азайды. Студент теориядан тәжірибеге көшті.

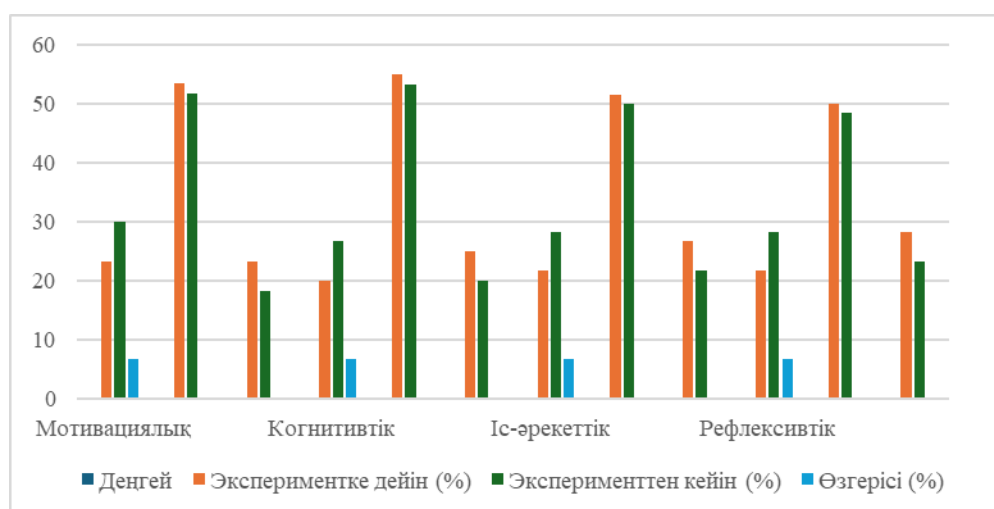
Ең үлкен тәуекел іс-әрекеттік блоктан күтілген еді. Педагогикалық кеңес алдындағы қысым мен ата-аналар тарапынан туындайтын талаптарды еңсеруде жоғары нәтижеге шыққан мамандар үлесі 20,0%-дан 55,0%-ға дейін көтерілді. Нәтиже күткеннен асып түсті. Тапсырмаларды орындауда қиналған орта деңгей 53,3%-дан 36,7%-ға шегерілсе, басқарушылық шешім қабылдай алмайтын дәрменсіз топ 26,7%-дан 8,3%-ға дейін қысқарды. Маман шешім қабылдауды үйренді.

Өз қателігіне есеп бере алмаған басшы ұжымды тоқырауға тірейді. Рефлексивтік критерий бойынша өзінің кәсіби қадамына сыни баға бере алатын жоғары көрсеткіштегі білім алушылар үлесі 23,3%-дан 56,7%-ға дейін ұлғайды. Орта деңгей 51,7%-да тұрақтады. Өзіне талдау жасай алмайтын төмен деңгейдегі респонденттер көрсеткіші 28,3%-дан 6,6%-ға дейін құлдырады. Барлық төрт компонент бойынша орын алған динамика мен сандық өзгерістердің арақатынасы көрнекі түрде бейнеленген (1-сурет). Модель өз тиімділігін толық дәлелдеді.



Сурет 1 - Эксперименттік топтың мектеп топ-менеджерінің сауаттылығы бойынша көрсеткіштері

Дәстүрлі оқу бағдарламасымен шектелген бақылау тобында күтілгендей ешқандай ілгерілеу байқалмады. Өзгеріс тым мардымсыз. Төрт критерий бойынша жоғары нәтиже көрсеткен студенттер үлесі орта есеппен бар болғаны 5-8% аралығында ғана ауытқып, қалыпты қателік шегінен аса алмады. Бұл тоқырауды білдіреді. Төмен деңгейдегі көрсеткіштердің азаю динамикасы да статистикалық тұрғыдан маңызды мән көрсетпей, алғашқы диагностикамен салыстырғанда бұрынғы күйінде қалды. Мұндай нәтиже қолданбалы модельсіз тек дәріс оқудың көшбасшылық қалыптастыруға қауқарсыз екенін айқындап берді (2-сурет).



Сурет 2 - Бақылау тобының мектеп топ-менеджерінің сауаттылығы бойынша көрсеткіштері

Эксперимент соңындағы бақылау кезеңі қос топтың аражігін ашық көрсетті. Статистикалық өңдеу екі топтың қорытынды нәтижелері арасында кездейсоқ емес, нақты сенімді айырмашылық барын айқындады ($p < 0,05$). Айырмашылық айқын. Эксперименттік аудиториядағы студенттердің 91,7%-ы басқарушылық шешім қабылдауға сенімді болса, 88,3%-ы командалық әрекетте өзін еркін сезіне бастаған. Мұндай ілгерілеу көшбасшылық қабілеттер бойынша 90,0%-ды, ал мектепті басқаруға деген қызығушылық бойынша 93,3%-ды құрады. Сандар нақты сөйлейді. Тұтас интегралды өлшемге салғанда, эксперименттік топ студенттерінің мектеп топ-менеджері сауаттылығы бастапқы шектен 31,8%-ға бірден өссе, дәстүрлі бағдарламамен жүрген бақылау тобында бұл өсім 7,4%-дан аса алмады.

Бұл сандық өзгеріс кездейсоқтық емес. Мотивациялық, когнитивтік, іс-әрекеттік және рефлексивтік төрт құрылымдық компоненттің қатар қозғалуы университет бағдарламаларына басқарушылық мазмұнды мақсатты түрде енгізудің қажеттігін негіздейді [22]. Дәстүрлі оқыту тар шеңберде қалды. Қазіргі білім кеңістігінде мұғалім сабақ берумен шектелмей, мектептің стратегиялық дамуы мен ұйымдастыру үдерістеріне араласатын толыққанды көшбасшы болуы шарт.

Тұжырымдар халықаралық ғылыми тәжірибемен үндеседі. Соңғы зерттеулерде білім беру ұйымдарының тиімділігі мұғалімнің көшбасшылық әлеуеті мен кәсіби шешім қабылдау қабілетіне тікелей байланысты екені дәйектелген. Тәжірибеге бағытталған модель осы жүйені реттейтін негізгі тетік ретінде қарастырылады.

Жобалық оқыту, кейс-технологиялар, іскерлік ойындар мен рефлексивтік күнделіктерді тұрақты қолданғанда ғана студенттер нақты әрекет етуге бейімделді [23]. Тәжірибесіз нәтиже жоқ. Әсіресе цифрлық білім беру ресурстарын басқарушылық ақпаратты талдау, жоспарлау және жедел коммуникация орнату үшін пайдалану жас мамандардың ұйымдастырушылық қысымды еңсеруіне оң әсер етті.

Зерттеудің қолданбалы мәні осы шешімдерді педагогикалық жоғары оқу орындарының білім беру бағдарламаларына енгізу мүмкіндігімен өлшенеді. Өзірленген әдістемелік нұсқаулықтар мен бағалау құралдары мектептердегі басқарушылық кадр сапасын арттыруға нақты ғылыми тірек береді [24]. Шешілмеген тұстар да бар.

Нәтижелерді мінсіз деуге әлі ерте:

- Зерттеу бірқатар педагогикалық жоғары оқу орындарының студенттері арасында ғана жүргізілді, сондықтан алынған көрсеткіштерді еліміздің барлық ЖОО-ларына толықтай жалпылау қосымша зерттеулерді қажет етеді;

- Басқарушылық сауаттылық қысқа мерзімді эксперимент шегінде бағаланды, ал оның кәсіби ортадағы ұзақ мерзімді тұрақтылығы әлі де белгісіз;

- Іріктеуге тек педагогикалық мамандық студенттері тартылды, яғни әртүрлі өңірлер мен бағдарламалар арасындағы салыстырмалы деректерді тереңірек талдау қажеттілігі сақталып отыр.

Болашақта мектеп топ-менеджерінің сауаттылығын цифрлық білім беру технологияларымен, жасанды интеллект құралдарымен және деректерге негізделген басқару тәсілдерімен тоғыстыру маңызды қадам болмақ. Басқару дағдыларының жас маманның кәсіби бейімделуі мен оқушылардың білім сапасына ықпалын бірнеше жылдық мониторинг арқылы бақылау ғылыми тұрғыдан аса қажет [25]. Бұл тәсіл педагогикалық білім мазмұнын жаңғыртып, болашақ кадрларды басқарушылық қызметке мақсатты даярлаудың әдістемелік негізін біржола бекітуге жол ашады.

Қорытынды

Педагогикалық жоғары оқу орындарының қабырғасында мұғалімді тек сабақ беруші шеңберінде даярлау үрдісі бүгінгі мектеп талабымен мүлде қабыспайды. Зерттеу мектеп топ-менеджерінің сауаттылығын кәсіби даярлықтың ажырамас бөлігі ретінде қалыптастырудың теориялық әрі қолданбалы қажеттілігін толық дәйектеп берді. Бұл сауаттылық - жай ғана құжатпен жұмыс істеу немесе бұйрықтарға бағыну емес. Оның іргетасы ұжымды ортақ мақсатқа жұмылдыратын мотивациялық, нормативтік-қаржылық тетіктерді білуге негізделген когнитивтік, нақты шешімдер қабылдайтын іс-әрекеттік және қателіктермен жұмыс істеуге мүмкіндік беретін рефлексивтік төрт компоненттің өзара үйлесімінен құралатыны ғылыми тұрғыда негізделді. Осы компоненттерді кешенді кіріктіретін авторлық педагогикалық модель сынақтан өтіп, өзінің өміршеңдігін дәлелдеді.

Күнделікті дәрістер көшбасшылық қасиетті қалыптастыра алмайды. Бұл - айқын факт. Оқу үдерісіне нақты мектептік кейстерді талдау, басқарушылық рөлдік симуляциялар, жобалық әдістер және рефлексивтік күнделіктерді жүйелі енгізу студенттердің шешім қабылдау батылдығын күрт өсірді. Эксперимент нәтижелері көрсеткендей, дәстүрлі

бағдарламамен жүрген бақылау тобындағы мардымсыз өзгерістермен (7,4%) салыстырғанда, эксперименттік топта интегралдық көрсеткіштің 31,8%-ға бірден артуы басқарушылық құзыреттерді аудиториялық ортада мақсатты түрде қалыптастыруға толық болатынын айқындады. Статистикалық мәліметтер нақты нәтижені көрсетіп отыр.

Зерттеу гипотезасы эмпирикалық деңгейде толық расталды. Егер университет бағдарламаларына мектептің стратегиялық дамуын жоспарлау, педагогикалық кеңесті ұйымдастыру және ата-аналармен дағдарыстық қарым-қатынасты реттеу сияқты тәжірибеге бағытталған шарттар кіріктірілсе, жас маман мектепке құр орындаушы емес, ұжымдағы өзгерістерді басқара алатын толыққанды топ-менеджер болып барады. Болжам өз жемісін берді.

Зерттеудің ғылыми құндылығы - отандық педагогикалық білім беру жүйесінде осы уақытқа дейін нақты жіктелмеген топ-менеджер сауаттылығының құрылымы мен өлшемшарттарын жүйелеуінде. Бұл теориялық олқылықтың орнын толтырады. Ал оның практикалық салмағы ұсынылған технологиялық кешеннің кез келген педагогикалық ЖОО білім беру бағдарламасына тікелей енгізуге дайын тұрғанымен өлшенеді. Әзірленген әдістемелік нұсқаулықтар мен бағалау құралдары білім беру ұйымдарындағы басқарушы кадр тапшылығын азайтуға ғана емес, мектептердегі реформалардың төменгі буында тұралап қалмауына тікелей ғылыми-тәжірибелік негіз ұсынады.

Әдебиеттер тізімі

1. Қазақстан Республикасының 2007 жылғы 27 шілдедегі № 319 «Білім туралы» Заңы (өзгерістер мен толықтырулармен). – Астана, 2007.
2. Қазақстан Республикасы Оқу-ағарту министрінің 2025 жылғы 24 ақпандағы № 31 бұйрығы. «Білім беру ұйымдарының педагогтеріне арналған кәсіптік стандарттарды бекіту туралы». – Астана, 2025.
3. Баймолдаев Т. М. Педагогический менеджмент и современное управление школой: монография. – Алматы: Ғылым, 2001. – 136 с.
4. Бахишева С. М. Педагогикалық жобалау: теориясы мен технологиясы: оқулық. – Алматы: Дәуір, 2011. – 336 б.
5. Әлқожаева Н. С., Жұмабекова Қ. Б. Педагогикалық менеджмент: оқу құралы. – 2-бас. – Алматы: Қазақ университеті, 2018. – 132 б.
6. Құрманалина Ш. Х., Бахишева С. М., Дәулетқалиева Н. Н., Архыматаева Н. Ж. Педагогикалық менеджмент: оқу құралы. – Астана: «Кәсіпқор» холдингі, 2018. – 131 б.
7. Drucker P. F. Management: Tasks, Responsibilities, Practices. New York: HarperBusiness, 1993
8. Fullan M. Leading in a Culture of Change. 2nd ed. Jossey-Bass, 2020
9. Mintzberg H. Understanding Organizations... Finally!: Structuring in Sevens. Berrett-Koehler Publishers, 2023
10. Darling-Hammond L. The Right to Learn: A Blueprint for Creating Schools That Work. San Francisco, CA: Jossey-Bass, 1997
11. Hattie J. Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement. Routledge, 2023. <https://doi.org/10.4324/9781003380542>
12. Robinson V. M. J. Student-Centered Leadership. San Francisco, CA: Jossey-Bass, 2011
13. Ng D. F. S. Instructional Leadership // Instructional Leadership and Leadership for Learning in Schools: Understanding Theories of Leading / ed. by T. Townsend. Palgrave Macmillan, 2019. P. 15–48. https://doi.org/10.1007/978-3-030-23736-3_2
14. Leithwood K., Sun J., Schumacker R. How School Leadership Influences Student Learning: A Test of “The Four Paths Model” // Educational Administration Quarterly. 2020. Vol. 56, no. 4. P. 570–599. <https://doi.org/10.1177/0013161X19878772>
15. UNESCO. Global Education Monitoring Report 2024/5: Leadership in Education – Lead for Learning. Paris: UNESCO, 2024. <https://doi.org/10.54676/EFLH5184>
16. OECD. Education Policy Outlook 2023: Empowering All Learners to Go Green. Paris: OECD Publishing, 2023. DOI: 10.1787/f5063653-en

17. Hallinger P., Liu S., Aung P. N., Yan M. A Systematic Review of Instructional Leadership Research Conducted with the Principal Instructional Management Rating Scale (PIMRS), 1983–2024 // Educational Management Administration & Leadership. 2025. Advance online publication. <https://doi.org/10.1177/17411432251358575>
18. Harris A., Jones M., Ismail N. Teacher Leadership and School/System Change: Exploring the Contemporary Evidence // International Journal of Education Policy and Leadership. 2025. Vol. 21, no. 3. P. 20. <https://doi.org/10.22230/ijep.2025v21n3a1589>
19. Huong V. T. M., Chau D. T. M., Quy T. T. Global Research Trends of Instructional Leadership: A Bibliometric Analysis From 2002 to 2023 // European Journal of Educational Management. 2025. Vol. 8, no. 2. P. 91–104. <https://doi.org/10.12973/eujem.8.2.91>
20. Sariakin S., Yeni M., Usman M. B., Mare A. S., Munzir M., Saleh M. Fostering a Productive Educational Environment: The Roles of Leadership, Management Practices, and Teacher Motivation // Frontiers in Education. 2025. Vol. 10. Article 1499064. <https://doi.org/10.3389/feduc.2025.1499064>
21. World Bank. Preparing and Supporting School Leaders. Washington, DC: World Bank, 2021
22. Zepeda S. J. Instructional Supervision: Applying Tools and Concepts. 5th ed. Routledge, 2025. <https://doi.org/10.4324/9781032642130>
23. Hallinger P., Kovačević J. Mapping the Intellectual Lineage of Educational Management, Administration and Leadership, 1972–2020 // Educational Management Administration & Leadership. 2022. Vol. 50, no. 2. P. 192–216. <https://doi.org/10.1177/17411432211006093>
24. Gümüş S., Bellibaş M. Ş., Esen M., Gümüş E. A Systematic Review of Studies on Leadership Models in Educational Research from 1980 to 2014 // Educational Management Administration & Leadership. 2018. Vol. 46, no. 1. P. 25–48. <https://doi.org/10.1177/1741143216659296>
25. Bush T. Theories of Educational Leadership and Management. 5th ed. London: SAGE Publications, 2020

A.B. Yertileuova, R. Želvys

FORMATION OF SCHOOL TOP MANAGEMENT LITERACY IN THE PROFESSIONAL TRAINING OF FUTURE TEACHERS

This study examines the development of school top-management literacy in the professional preparation of future teachers. The increasing complexity of management processes in contemporary education, the need to enhance the competitiveness of educational institutions, and new professional requirements for teaching staff have intensified the importance of developing managerial competencies among future teachers. The purpose of the study is to identify the theoretical and methodological foundations for developing school top-management literacy in the professional preparation of future teachers, systematize its structural components, and substantiate the pedagogical opportunities for its development in higher education. Within the framework of the study, school top-management literacy is conceptualized as an interconnected system of competencies required for the effective management of an educational institution, including strategic thinking, leadership, managerial decision-making, communication culture, teamwork, change management, and organizational planning. While national curricula confine the teacher to a narrow subject-methodological scope, international experience assigns them the role of an accountable manager who directly shapes the strategic decisions of the entire institution. Consequently, concepts in educational management and leadership were compared, and the functional profile of a school top manager was directly linked to the teacher's professional training level. That is the core issue. Managerial literacy cannot be measured solely by administrative instructions. The analysis indicates that the framework rests on six specific competencies: strategic vision, decision-making, team synergy, and organizational skills. Training only an instructor will not bridge the managerial vacuum. Therefore, the study substantiates the need to integrate educational management modules into university programs and to expand practical cases grounded in real school conflicts and resource risks. A professional must make decisions. Qualitative progress remains unattainable without preparing teachers to participate in organizational governance, while the proposed conclusions provide a scientific-methodological foundation for modernizing higher education programs.

Keywords: future teachers, professional preparation, school top management, managerial literacy, managerial competence, leadership, educational management.

А.Б. Ертілеуова, Р. Желвис

ФОРМИРОВАНИЕ ГРАМОТНОСТИ ШКОЛЬНОГО ТОП-МЕНЕДЖЕРА В ПРОФЕССИОНАЛЬНОЙ ПОДГОТОВКЕ БУДУЩИХ ПЕДАГОГОВ

В исследовании рассматривается проблема формирования грамотности в области топ-менеджмента школы в процессе профессиональной подготовки будущих педагогов. Усложнение управленческой деятельности в современной системе образования, необходимость повышения конкурентоспособности образовательных организаций и новые профессиональные требования к педагогическим кадрам усиливают актуальность развития управленческих компетенций будущих педагогов. Цель исследования - определить теоретические и методологические основы формирования грамотности в области топ-менеджмента школы в процессе профессиональной подготовки будущих педагогов, систематизировать ее структурные компоненты и обосновать педагогические возможности ее развития в условиях высшего образования. В рамках исследования грамотность в области топ-менеджмента школы рассматривается как взаимосвязанная система компетенций, необходимых для эффективного управления образовательной организацией, включающая стратегическое мышление, лидерство, принятие управленческих решений, коммуникативную культуру, командное взаимодействие, управление изменениями и планирование организационной деятельности. Если отечественные учебные программы удерживают учителя в узких предметно-методических рамках, то зарубежный опыт возлагает на него роль ответственного менеджера, напрямую влияющего на стратегические решения всей организации. Поэтому концепции в области образовательного менеджмента и лидерства были сопоставлены, а функциональная структура школьного топ-менеджера напрямую связана с уровнем профессиональной подготовки педагога. Это и есть ключевая проблема. Управленческая грамотность не измеряется одними лишь кабинетными инструкциями. Как показывает анализ, опора системы сводится к шести конкретным компетенциям: стратегическому видению, принятию решений, командной согласованности и организационным навыкам. Подготовка простого предметника не способна заполнить управленческий вакуум. В связи с этим обоснована необходимость внедрения модулей образовательного менеджмента в университетские программы и расширения практических кейсов, основанных на реальных школьных конфликтах и ресурсных рисках. Специалист должен принимать решения. Качественный прорыв невозможен без вывода педагога на уровень непосредственного участия в процессах управления организацией, а предложенные выводы служат научно-методической опорой для модернизации вузовских программ.

Ключевые слова: будущий педагог, профессиональная подготовка, топ-менеджмент школы, управленческая грамотность, управленческая компетентность, лидерство, образовательный менеджмент.

References

1. Law of the Republic of Kazakhstan No. 319 dated July 27, 2007, “*Bilim týraly*” [On Education] (with amendments and additions). (2007). Astana.
2. Order of the Minister of Education of the Republic of Kazakhstan No. 31 dated February 24, 2025, “*Bilim berý uıymdarynyń pedagogterine arnalǵan kásiptik standarttardy bekítý týraly*” [On approval of professional standards for teachers of educational organizations]. (2025). Astana.
3. Baimoldayev, T. M. (2001) *Pedagogicheskiy menedzhment i sovremennoe upravlenie shkoloy* [Pedagogical management and modern school management]: monografiya. Gylym.
4. Bakhisheva, S. M. (2011). *Pedagogikalyq jobalau: teoriyasy men texnologiyasy* [Pedagogical design: theory and technology]. Daur.
5. Alkozhaeva, N. S., & Zhumabekova, K. B. (2018). *Pedagogikalyq management: oqu quraly* [Pedagogical management] (2nd ed.). Kazakh University.
6. Kurmanalina, Sh. Kh., Bakhisheva, S. M., Dauletkaliyeva, N. N., & Arkhymatayeva, N. Zh. (2018). *Pedagogikalyq menedjment, oqu quraly* [Pedagogical management]. Kasipkor Holding.
7. Drucker, P. F. (1993). *Management: Tasks, responsibilities, practices*. HarperBusiness.
8. Fullan, M. (2020). *Leading in a culture of change* (2nd ed.). Jossey-Bass.
9. Mintzberg, H. (2023). *Understanding organizations... finally!: Structuring in sevens*. Berrett-Koehler Publishers.

10. Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. Jossey-Bass.
11. Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9781003380542>
12. Robinson, V. M. J. (2011). *Student-centered leadership*. Jossey-Bass.
13. Ng, D. F. S. (2019). Instructional leadership. In T. Townsend (Ed.), *Instructional leadership and leadership for learning in schools: Understanding theories of leading* (pp. 15–48). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-23736-3_2
14. Leithwood, K., Sun, J., & Schumacker, R. (2020). How school leadership influences student learning: A test of “The Four Paths Model”. *Educational Administration Quarterly*, 56(4), 570–599. <https://doi.org/10.1177/0013161X19878772>
15. UNESCO. (2024). *Global education monitoring report 2024/5: Leadership in education – Lead for learning*. UNESCO. <https://doi.org/10.54676/EFLH5184>
16. OECD. (2023). *Education policy outlook 2023: Empowering all learners to go green*. OECD Publishing. <https://doi.org/10.1787/f5063653-en>
17. Hallinger, P., Liu, S., Aung, P. N., & Yan, M. (2025). A systematic review of instructional leadership research conducted with the Principal Instructional Management Rating Scale (PIMRS), 1983–2024. *Educational Management Administration & Leadership*, Advance online publication. <https://doi.org/10.1177/17411432251358575>
18. Harris, A., Jones, M., & Ismail, N. (2025). Teacher leadership and school/system change: Exploring the contemporary evidence. *International Journal of Education Policy and Leadership*, 21(3), 1–20. <https://doi.org/10.22230/ijep.2025v21n3a1589>
19. Huong, V. T. M., Chau, D. T. M., & Quy, T. T. (2025). Global research trends of instructional leadership: A bibliometric analysis from 2002 to 2023. *European Journal of Educational Management*, 8(2), 91–104. <https://doi.org/10.12973/eujem.8.2.91>
20. Sariakin, S., Yeni, M., Usman, M. B., Mare, A. S., Munzir, M., & Saleh, M. (2025). Fostering a productive educational environment: The roles of leadership, management practices, and teacher motivation. *Frontiers in Education*, 10, Article 1499064. <https://doi.org/10.3389/educ.2025.1499064>
21. World Bank. (2021). *Preparing and supporting school leaders*. World Bank.
22. Zepeda, S. J. (2025). *Instructional supervision: Applying tools and concepts* (5th ed.). Routledge. <https://doi.org/10.4324/9781032642130>
23. Hallinger, P., & Kovačević, J. (2022). Mapping the intellectual lineage of educational management, administration and leadership, 1972–2020. *Educational Management Administration & Leadership*, 50(2), 192–216. <https://doi.org/10.1177/17411432211006093>
24. Gümüş, S., Bellibaş, M. Ş., Esen, M., & Gümüş, E. (2018). A systematic review of studies on leadership models in educational research from 1980 to 2014. *Educational Management Administration & Leadership*, 46(1), 25–48. <https://doi.org/10.1177/1741143216659296>
25. Bush, T. (2020). *Theories of educational leadership and management* (5th ed.). SAGE Publications

Авторлар туралы мәлімет:

Ертілеуова Аида Балқашбайқызы - PhD докторант, Yessenov University, электрондық пошта: e-mail: yertileuovaaida@gmail.com

Желвис Римантас - профессор, Вильнюс университеті, Философия факультеті, Білім беру ғылымдары институты, Вильнюс қ., Литва, e-mail: rimantas.zelvys@fsf.vu.lt

Сведения об авторе:

Ертілеуова Аида Балқашбайқызы - PhD-докторант, Yessenov University, электронная почта: yertileuovaaida@gmail.com

Желвис Римантас - профессор, Вильнюсский университет, Философский факультет, Институт образовательных наук, г. Вильнюс, Литва, электронная почта: rimantas.zelvys@fsf.vu.lt

Information about author:

Yertileuova Aida Balkhashbaykyzy - PhD doctoral student, Yessenov University, e-mail: yertileuovaaida@gmail.com

Zelvys Rimantas - Professor, Vilnius University, Faculty of Philosophy, Institute of Educational Sciences, Vilnius, Lithuania, e-mail: rimantas.zelvys@fsf.vu.lt

Ә.Қ. Құлмағанбет^{1*}, Д.С. Оспанов², Т.С. Ерсұлтанова³^{1,2,3}ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана қ., Қазақстан Республикасы

*e-mail: asemkulmaganbetova@mail.ru

¹ORCID 0009-0006-0774-3538, ²ORCID 0009-0006-4418-0770, ³ORCID 0000-0003-3286-9096

ПЕДАГОГИКА ЖӘНЕ ПСИХОЛОГИЯ БОЙЫНША ТЕСТІЛЕУДІ ПСИХОМЕТРИЯЛЫҚ ТАЛДАУ: ТЕХНИКАЛЫҚ ЖӘНЕ КӘСІПТІК БІЛІМ БЕРУ ЖҮЙЕСІНДЕГІ БАҒАЛАУ САПАСЫНЫҢ КЕПІЛІ

Бұл зерттеу Индустрия 4.0 жаһандық сын-қатерлері мен білім беруді цифрландыру жағдайында техникалық және кәсіптік білім беру (ТЖКББ) ұйымдары педагогтерінің кәсіби құзыреттілігін бағалау сапасын арттыру мәселелеріне арналған. Қазақстан Республикасы Оқу-ағарту министрлігінің 2025 жылғы 25 ақпандағы №32 бұйрығы аясындағы реформалар контекстінде «Педагогика» және «Психология» пәндері бойынша тестілеу құралдарын психометриялық талдаудың (макро және микродеңгейде) ғылыми-әдістемелік маңызы негізделген. Зерттеу барысында Классикалық тест теориясының (СТТ) негізгі көрсеткіштері - Кронбах альфасы (α), валидтілік, тапсырманың қиындық деңгейі (P), саралау коэффициенті (D) және дистракторлардың тиімділігі бойынша 280 қатысушы мен 50 тест тапсырмасының тестілеу нәтижелеріне математикалық-статистикалық талдау жүргізілді. Германия, Ұлыбритания, Сингапур және АҚШ-тың халықаралық бағалау модельдерімен салыстырмалы талдау жасалып, отандық бағалау жүйесін модернизациялау бағыттары айқындалды. Психометрикалық мониторинг тест сапасын арттыруға, екіұшты сұрақтар мен жұмыс істемейтін дистракторларды жоюда жоғары тиімділік көрсетті. Зерттеу нәтижелері тест тапсырмалары банкін сапалы жаңартуға, педагогтердің «профильдік кемшіліктерін» анықтауға және біліктілікті арттыру курстарын мақсатты жоспарлауға мүмкіндік береді. Осылайша, психометриялық сүзгі ТЖКББ жүйесіндегі бағалаудың объективтілігі мен академиялық адалдығының негізгі кепілі ретінде дәлелденді.

Түйін сөздер: ТЖКББ жүйесі, психометриялық талдау, Кронбах альфасы, қиындық индексі, дистракторлық талдау, бағалау сапасы.

Кіріспе

Педагогика мен психология өзара тығыз байланысты және зерттеу өзектілігі. ХХІ ғасырдың дамыған елдері индустриалды және ғылыми-техникалық кезеңін басынан өткерген, экономикада, мәдениетте және саяси-әлеуметтік құрылымда едәуір жоғары деңгейге көтерілген 4.0 индустриясы бар бәсекеге қабілетті ғылыммен және біліммен сипатталады. 4.0 индустриясында экономикалық өсу табиғи ресурстарға емес, инновацияларға және бәсекеге қабілетті адами капиталға негізделген.

Бұл трансформация білім берудің алдына күрделі міндеттер қояды. Индустрия 4.0 жаһандық сын-қатерлері мен экономиканы цифрландыру жағдайында колледждердегі мамандарды даярлау сапасы оқытушылар құрамының кәсіби құзыреттілігімен тікелей байланысты. Бұл тұрғыда педагогтерді бағалау және аттестациялау енді тек әкімшілік рәсімдер емес, адами капиталды басқару және білім беру үдерісінің сапасын қамтамасыз ету құралына айналуға [1].

Қазіргі таңда техникалық және кәсіптік білім беру (ТЖКББ) жүйесінде еңбек нарығының талаптарына сай бәсекеге қабілетті мамандар даярлау басты міндеттердің бірі болып табылады. Әсіресе, гуманитарлық және педагогикалық бағыттағы мамандықтар үшін «Педагогика» және «Психология» пәндері бойынша білім сапасын тексеру - студенттердің кәсіби даярлығының іргетасы. Дегенмен, білім беру тәжірибесінде қолданылатын тестілеу құралдарының сапасы әрқашан талапқа сай келе бермейді. Көбіне тест сұрақтары тек жаттанды ақпаратты тексеруге бағыттталып, сыни ойлау мен кәсіби жағдаяттарды талдау

деңгейін қамти алмайды. Бұл мәселені шешудің жалғыз ғылыми жолы - психометриялық талдау (тесттердің сапасын математикалық және статистикалық әдістермен бағалау) [2].

Техникалық және кәсіптік білім беру ұйымдарының педагогтерін аттестаттау мәселелерін реттейтін құқықтық база бірқатар нормативтік құқықтық актілерге негізделген, олардың ішінде ең маңыздысы Қазақстан Республикасы Білім министрінің 2016 жылғы 27 қаңтардағы № 83 бұйрығы (Қазақстан Республикасы Оқу-ағарту министрінің 25 ақпандағы №32 бұйрығымен енгізілген өзгерістер мен толықтырулар) болып табылады.

Педагогика мен психология өзара тығыз байланысты, абстрактілі әрі контекстке тәуелді ғылымдар болғандықтан, олар бойынша құрастырылған тесттер ерекше икемділікті талап етеді.

Қазіргі заманғы сертификаттау жүйесінің философиясы бағдарламалық құжаттарға сәйкес әлеуметтік әділеттілік, академиялық адалдық және объективтілік принциптеріне негізделген. «Педагогтерді аттестаттаудан өткізу қағидалары мен шарттарын бекіту туралы» Қазақстан Республикасы Оқу-ағарту министрлігінің 2025 жылғы 25 ақпандағы №32 бұйрығымен елеулі өзгерістер енгізілді. Қазақстан Республикасы Әділет министрлігінде тіркелген және Педагогикалық қызметкерлерді аттестациялауды өткізу ережелері мен шарттарын қайта қарайтын акт аттестаттау тәртібі мен критерийлерін жаңартады [3].

Зерттеу материалдары мен әдістері

Бұл зерттеу сандық әдіснамаға негізделген тест тапсырмаларының психометриялық сипаттарын талдауға бағытталған. Оның басты мақсаты - субъектілерді бағалауда тесттің «өлшеуіш құрал» ретіндегі сапасын, дәлдігін және қатесіздігін дәлелдеу. Психометриялық талдау тесттің сапасын екі деңгейде айқындайды: тұтас тесттің сапасы (макродеңгей) және әрбір жеке тапсырманың сапасы (микродеңгей).

1. Сенімділік (Reliability) және Кронбах альфасы (α)

Сенімділік - тест нәтижелерінің тұрақтылығы мен дәлдігін көрсетеді. Ішкі үйлесімділікті анықтау үшін психометрияда Кронбах альфасы қолданылады:

$$\alpha = k - 1/k (1 - \sigma^2 X^2 \sum_{i=1}^k \sigma_i^2)$$

Мұндағы: k - тесттегі тапсырмалардың жалпы саны;

σ_i^2 - әрбір жеке i -ші тапсырманың дисперсиясы;

$\sigma^2 X^2$ - тесттің жалпы балының дисперсиясы.

«Педагогтерді аттестаттаудан өткізу қағидалары мен шарттарын бекіту туралы» Қазақстан Республикасы Оқу-ағарту министрлігінің 2025 жылғы 25 ақпандағы №32 бұйрық талаптарына сәйкес, ресми сертификаттауға қолданылатын тесттердің сенімділік коэффициенті $\alpha \geq 0.80$ болуы шарт. Көрсеткіш төмен болса, тест тапсырмалары қайта қаралады.

2. Валидтілік (Validity)

Тесттің дәл өзі өлшеуге тиісті құзыреттілікті (мысалы, педагогикалық шеберлікті немесе психологиялық білімді) қаншалықты дәл өлшейтіндігі. Жаңа реформа аясында мазмұндық валидтілік (тест тапсырмаларының оқу бағдарламасына сәйкестігі) және конструктілік валидтілік (болашақ маманның кәсіби бейінін аша алуы) басты назарға алынады.

3. Тапсырманың қиындық деңгейі (Difficulty Index, P)

Бұл - берілген сұраққа дұрыс жауап берген сыналушылардың үлесі:

$P = NR$. Мұндағы R - сұраққа дұрыс жауап бергендер саны, N - тест тапсырғандардың жалпы саны. Оңтайлы көрсеткіш: $0.30 \leq P \leq 0.70$. $P > 0.85$ болса - сұрақ тым жеңіл, $P < 0.20$ болса - сұрақ тым қиын деп танылады.

4. Саралау қабілеті (Discrimination Index, D)

Тесттен ең жоғары балл алғандар (жоғары топ - P_H) мен төмен балл алғандардың (төменгі топ - P_L) сұраққа жауап беру айырмашылығы:

$D = P_H - P_L$. $D \geq 0.30$ - жоғары саралау қабілетінің белгісі. Егер $D < 0.20$ болса, бұл тест тапсырмасы күшті және әлсіз педагогтерді ажырата алмайды, демек жарамсыз [4].

Техникалық және кәсіптік білім беру ұйымдарының педагогтерін бағалаудың әлемдік модельдеріне арналған халықаралық салыстырмалы 1-кестені ұсынамыз.

Кесте 1 - Халықаралық тәжірибе

Ел/Модель	Бағалаудың негізгі философиясы	Бағалау құралдары мен әдістері	Ерекшелігі мен артықшылығы
Германия (<i>Дуальді білім беру моделі</i>)	Өндірістік және құзыреттілік икемділік. Мұғалім ең алдымен жоғары деңгейлі өндіріс шебері болуы тиіс	Өндірістік емтихандар; Дидактикалық кейс-стади;	Педагогтің өндірістік дағдылары мен теориялық білімі қатар бағаланады. Субъективтілік деңгейі өте төмен.
Ұлыбритания (Ұлттық кәсіби стандарттар NOS) National Occupational Standards)	Тәуелсіз аудиторлық бағалау. Нақты нәтиже мен жұмыс орнындағы тиімділікті өлшеу	Кәсіби портфолио (Portfolio of Evidence); • Жұмыс үстінде бақылау (Work-based observation)	Мұғалімдерді мемлекеттік шенеуніктер емес, тәуелсіз кәсіби қауымдастық өкілдері «соқыр бағалау» әдісімен тексереді.
Сингапур (<i>Нәтижелілік пен даму моделі-EPMS</i>) Enhanced Performance Management System	Үздіксіз кәсіби өсу және рефлексия. Бағалау - жазалау құралы емес, даму индикаторы.	Жылдық жоспарлау және өзін-өзі бағалау; • Көшбасшылық және инновациялық жобалардың нәтижесі	Жүйе мұғалімдерді үш бағыт бойынша бөледі: педагогикалық шеберлік, көшбасшылық немесе зерттеушілік.
АҚШ (Ұлттық кәсіби педагогикалық стандарттар-NBPTS) National Board for Professional Teaching Standards	Ерікті және жоғары деңгейлі сертификаттау. Академиялық адалдық пен терең кәсібилік.	Бейнежазбалар (сабақты талдау) • Пән бойынша компьютерлік тестілеу (ПББ сияқты); • Студенттердің жұмыстарын талдау (кейс шешу)	Сертификат алу өте қиын, бірақ ол мұғалімнің жалақысын бірден 10-15%-ға өсіреді және мәртебесін айқындайды.

«Халықаралық тәжірибені талдау көрсеткендей (Кесте-1), дамыған елдердің Техникалық және кәсіптік білім беру ұйымының жүйесінде педагогтерді бағалау қағазбастылықтан әлде қашан арылған және нақты нәтижеге бағытталған. Германияның дуальді жүйесі [5], Ұлыбританияның тәуелсіз портфолио аудиты [6] және Сингапурдың даму модельдері [7] заманауи сертификаттаудың негізін құрайды. «Педагогтерді аттестаттаудан өткізу қағидалары мен шарттарын бекіту туралы» Қазақстан Республикасы Оқу-ағарту министрлігінің 2025 жылғы 25 ақпандағы №32 бұйрығымен енгізілген реформалар дәл осы халықаралық тәжірибеге, атап айтқанда, бағалаудың ашықтығы мен цифрлық платформалар арқылы адам факторын азайтуға бағытталған. Бұл ретте, АҚШ пен Ұлыбритания тәжірибесіндегідей, тестілеу және портфолио бағалау құралдарының сапасын қамтамасыз ету үшін бізге отандық бағалау жүйесінде психометриялық талдаудың математикалық модельдерін енгізу өте маңызды.

Нәтижелер және талқылау

Психометриялық талдау нәтижелері тесттің жоғары сапасын көрсетті. Практикалық матрица Техникалық және кәсіптік білім беру жүйесінде «Педагогика және психология» пәндері бойынша өткізілген пилоттық тестілеу нәтижелеріне жасалған психометриялық талдаудың модельдік матрицасы төмендегі 2- кестеде көрсетілген.

Кесте 2 - Тест тапсырмаларының психометриялық сипаттамасы мен шешімдер матрицасы

Тапсырма №	Тест тақырыбы (Педагогика/ Психология)	Қиындық деңгейі (P)	Саралау коэффициенті (D)	Дистракторлардың күйі	Психометриялық қорытынды және шешім
№1	Дидактиканың негізгі принциптері	0.65	0.42	Барлық нұсқалар тиімді жұмыс істейді	Идеалды тапсырма. Тест базасында өзгеріссіз қалдырылады.
№2	Жас ерекшелік психологиясы (дағдарыстар)	0.15	0.08	Студенттердің 80%-ы бір қате нұсқаны таңдаған	Тым қиын әрі сараламайды. Сұрақ мәтіні екіұшты жазылған. Қайта өңдеуге жіберіледі.
№3	Педагогикалық қарым-қатынас стильдері	0.92	0.12	3 дистракторды ешкім таңдамаған	Тым жеңіл сұрақ. Саралау қуаты төмен. Дистракторларын күшейту керек немесе алып тастау қажет.
№4	Тұлғаның бағыттылығы мен мотивациясы	0.45	0.38	Оқушылардың таңдауы біркелкі бөлінген	Оңтайлы сұрақ. Үздіктер мен үлгерімі төмендерді жақсы ажыратады. Базада қалады.

Кесте 2 көріп отырғанымыздай, дистракторлық талдау арқылы жұмыс істемейтін қате жауаптарды анықтауға болады. Мысалы, №3 тапсырмадағы үш нұсқаны ешкім белгілемегендіктен, ол сұрақ өзінің құндылығын жойған.

Педагогика және психология тесттерін талдау ерекшеліктері.

Педагогика және психология пәндері бойынша тест құрастыруда психометрияның берер көмегі зор. Мысалы, «Конфликтология» немесе «Даму психологиясы» бөлімдері бойынша жағдаяттық (контекст) тест тапсырмасы әзірленгенде, дистракторларды талдау маңызды.

Дистракторлық талдау. Егер тест тапсырмасының төрт жауап нұсқасының біреуін (дұрыс емес нұсқаны) ешбір педагог таңдамаса, ол дистрактор «жұмыс істемейді» деген сөз. Демек, ол тесттің сапасын төмендетеді [8].

Психометриялық мониторинг нәтижесінде Техникалық және кәсіптік білім беру оқытушылары: Жарамсыз, екіұшты сұрақтарды базадан шығарады; Педагогтердің қай тақырыпты (мысалы, «Педагогикалық» немесе «Жас ерекшелік психологиясы») нашар меңгергенін нақты статистикамен көре алады;

Дистракторлық Талдау (Distractor Analysis) дегеніміз не?

Дистрактор (ағылш. *distract* - алаңдату, назарды басқа жаққа бұру) - тақырыпты терең білмейтін, тек үстірт оқыған педагогті өзіне тартуы тиіс қате жауап нұсқасы. Психометриялық ғылымда дистракторлардың жұмыс істеу сапасы былай анықталады:

Тиімді (Жұмыс істейтін) дистрактор: Тест тапсырған әлсіз немесе дайындығы орташа педагогтердің кем дегенде 5%-ы таңдаған нұсқа. Ол неғұрлым логикалық тұрғыдан дұрыс сияқты көрінсе, тесттің сапасы соғұрлым жоғары болады.

Тиімсіз (Жұмыс істемейтін) дистрактор: Ешбір педагог таңдамаған, тым күлкілі немесе контекстке мүлдем сәйкес келмейтін қате нұсқа.

Мәселе: Егер 4 жауап нұсқасы бар тесттің 2 дистракторын ешкім таңдамаса, бұл сұрақ өз құндылығын жояды. Ол автоматты түрде 2 нұсқалы жеңіл сұраққа айналады (дұрыс жауапты кездейсоқ табу мүмкіндігі 25%-дан бірден 50%-ға көтеріледі). Психометриялық мониторинг мұндай «өлі» дистракторларды бірден анықтап, оларды күшейтуді немесе ауыстыруды талап етеді [9, 10].

Тест нәтижелерін статистикалық (SPSS, Excel немесе арнайы Ұлттық платформа алгоритмдері) өңдеу арқылы Техникалық және кәсіптік білім беру жүйесінде мынадай үлкен 3 мәселе шешіледі:

а) жарамсыз және екіұшты сұрақтарды базадан шығару;

Педагогика мен психологияда терминдер көп (мысалы, Виготскийдің, Пиаженің немесе Блумның теориялары). Егер сұрақ анық қойылмаса, екі нұсқа теңдей дұрыс болуы мүмкін. Психометриядағы Саралау коэффициенті (D) арқылы бұл бірден көрінеді. Егер үздік педагогтер сұрақтан жаппай қателесіп, төмен үлгерімділер кездейсоқ белгілеп кетсе, тест тапсырмасының құрылымы қате. Мұндай екіұштытест тапсырмалары педагогтің тағдырына (санатына) нұқсан келтірмеуі үшін базадан автоматты түрде өшіріледі.

ә) педагогтердің «Профильдік кемшіліктерін» (Нақты статистиканы) көру. Психометриялық есеп беру жүйесі Техникалық және кәсіптік білім беру басшылығына педагогтердің қай тақырыпты нашар меңгергенін нақты пайызбен көрсетіп береді. *Мысалы:* Статистика көрсеткендей, педагогтердің 85%-ы «Дидактика» сұрақтарына мүлтіксіз жауап берген, бірақ «Жас ерекшелік психологиясы» немесе «Инклюзивті білім беру» бөліміне келгенде дұрыс жауап көрсеткіші 40%-ға түсіп кеткен. *Нәтиже:* Бұл ақпарат біліктілікті арттыру институттарына келесі жолы педагогтерге қандай бағытта курс ұйымдастыру керектігін нақты нұсқайды.

б) бағалаудың субъективтілігінен толықтай арылу;

Бұрын педагогтердің білімі мен біліктілігін колледж әкімшілігі, комиссия мүшелері өздерінің жеке көзқарасы мен қарым-қатынасына қарай бағалай алатын (субъективтілік). Жаңа жүйеде психометриялық сүзгіден өткен тесттерді компьютер тексереді. Адами фактор, «ұнату-ұнатпау» сезімдері рөл ойнамайды. Баға тек санмен және статистикамен дәлелденеді [11].

Осылайша, педагогика және психология пәндері бойынша дистракторлық талдау мен психометриялық мониторинг жүргізу - бағалаудың жай ғана құралы емес, бұл білім берудегі әділеттіліктің басты шарты. Ол Техникалық және кәсіптік білім беру оқытушыларының кәсіби деңгейін айнадай айқын көрсетіп, екіұштылықты жояды және «Педагогтерді аттестаттаудан өткізу қағидалары мен шарттарын бекіту туралы» Қазақстан Республикасы Оқу-ағарту министрлігінің 2025 жылғы 25 ақпандағы №32 бұйрығында бекітілген академиялық адалдық принципін толық қамтамасыз етеді.

Бағалау құралдарының математикалық дәлдігі: Психометриялық талдаудың негізгі көрсеткіштері - Кронбах альфасы ($\alpha 0.80$), валидтілік, тапсырманың қиындық деңгейі (P) және саралау коэффициенті (D) - тест тапсырмаларын тек макродеңгейде (тұтас тест) емес, микродеңгейде де (әрбір жеке сұрақ) сүзгіден өткізуге мүмкіндік береді. Пилоттық тестілеу нәтижелері көрсеткендей, көрсеткіштер екіұшты және тым жеңіл/қиын сұрақтарды дер кезінде анықтап, тест базасын үнемі модернизациялап отырудың жалғыз ғылыми жолы болып табылады [12].

«Педагогика және психология» пәні бойынша психометрикалық нәтижелері.

Зерттеудің мақсаты мен деректер құрылымы.

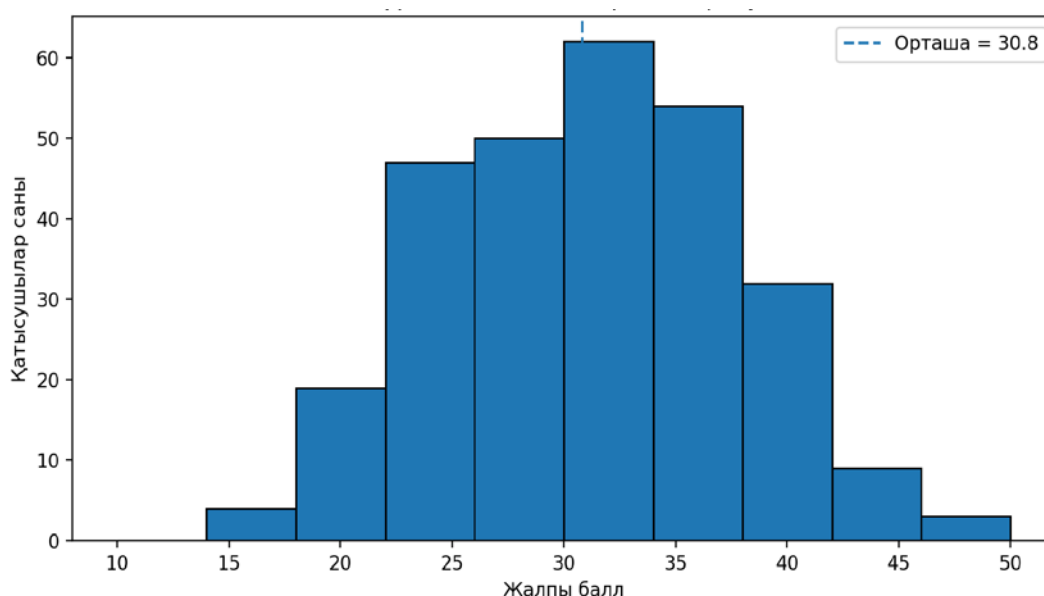
Талдаудың мақсаты - тесттің тұтас сенімділігін, әр тапсырманың қиындық және саралау қасиеттерін, тапсырма - тест байланысын, дистракторлардың тиімділігін және тақырыптар

бойынша меңгеру деңгейін бағалау. Классикалық тест теориясының көрсеткіштері қолданылды: Кронбах альфасы, P, D, RIT, RIR және pBis.

Кесте 3 - Тест бойынша жалпы статистикалық көрсеткіштер

Көрсеткіш	Нәтиже
Қатысушылар саны	280
Тапсырмалар саны	50
Орташа балл	30.81
Орташа орындалу	61.6%
Стандарттық ауытқу	6.58
Кронбах альфасы	0.778
Стандарттық өлшеу қатесі (SEM)	3.10
Минимум–максимум	15-46

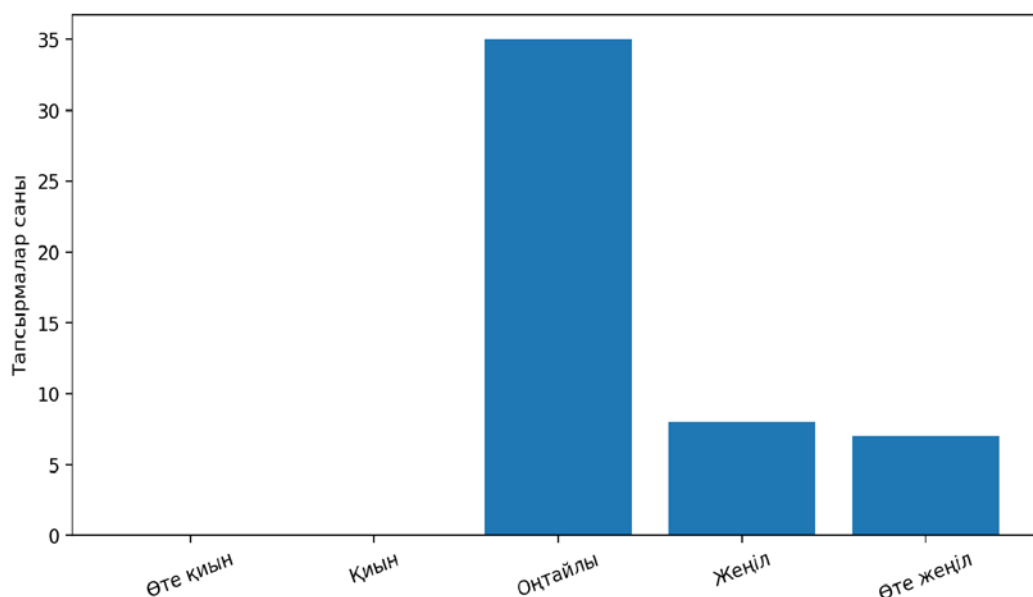
Кесте 3 бойынша тест бойынша жалпы статистикалық көрсеткіштері. Интерпретация. Орташа нәтиже 30.81 балл немесе 61.6% болды. Кронбах альфасы $\alpha=0.778$. Бұл тесттің ішкі үйлесімділігі қанағаттанарлық деңгейде екенін, алайда жоғары жауапкершілікті бағалауда көзделетін 0,80 межесіне кішкене жетпейтінін көрсетеді. SEM=3.10 балл болғандықтан, жеке қатысушы нәтижесін интерпретациялағанда шамамен ± 3.1 балл өлшеу қатесін ескеру қажет.



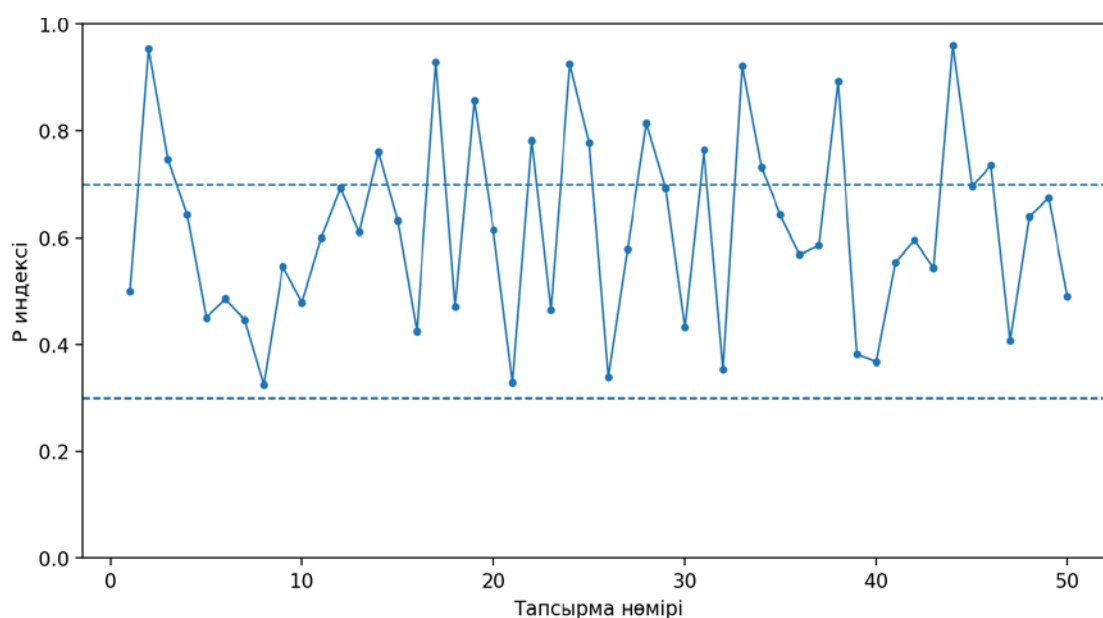
1 сурет - Жалпы балдардың таралуы

Тапсырмалардың қиындық деңгейіне тоқталатын болсақ, 2 суретте төмендегідей.

50 тапсырманың ішінде өте қиын – 0, қиын – 0, оңтайлы – 35, жеңіл – 8, өте жеңіл – 7 тапсырма анықталды. Р индексі 0,30-0,70 аралығындағы тапсырмалар тесттің негізгі ақпараттық бөлігін құрайды. Өте жеңіл немесе өте қиын тапсырмалар мазмұндық мақсатпен арнайы енгізілмесе, олардың саралау мүмкіндігі төмендеуі ықтимал.



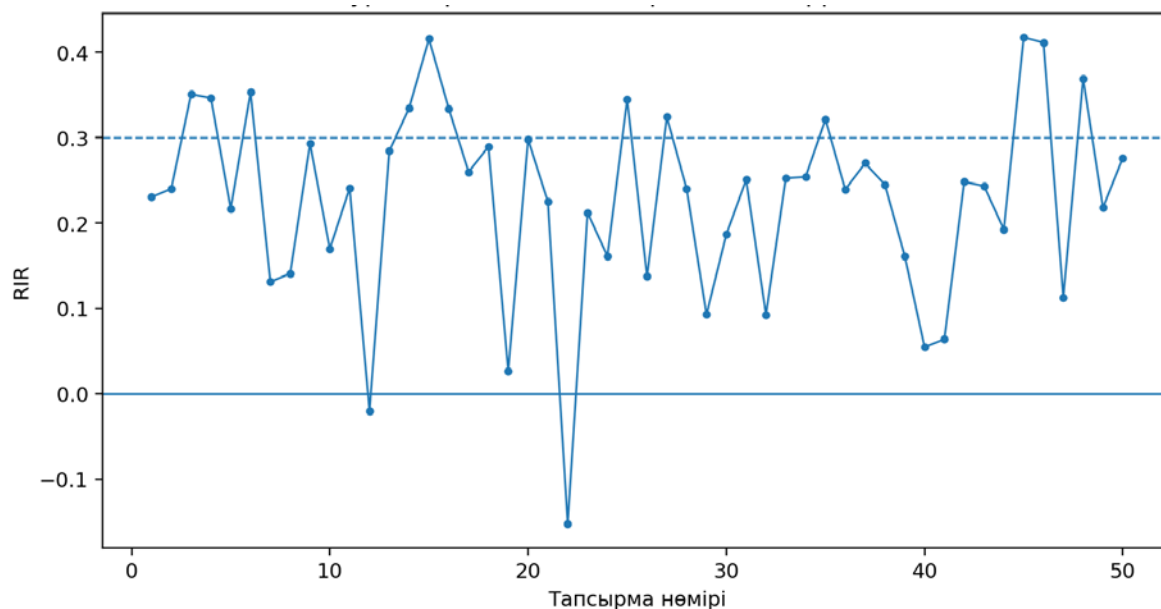
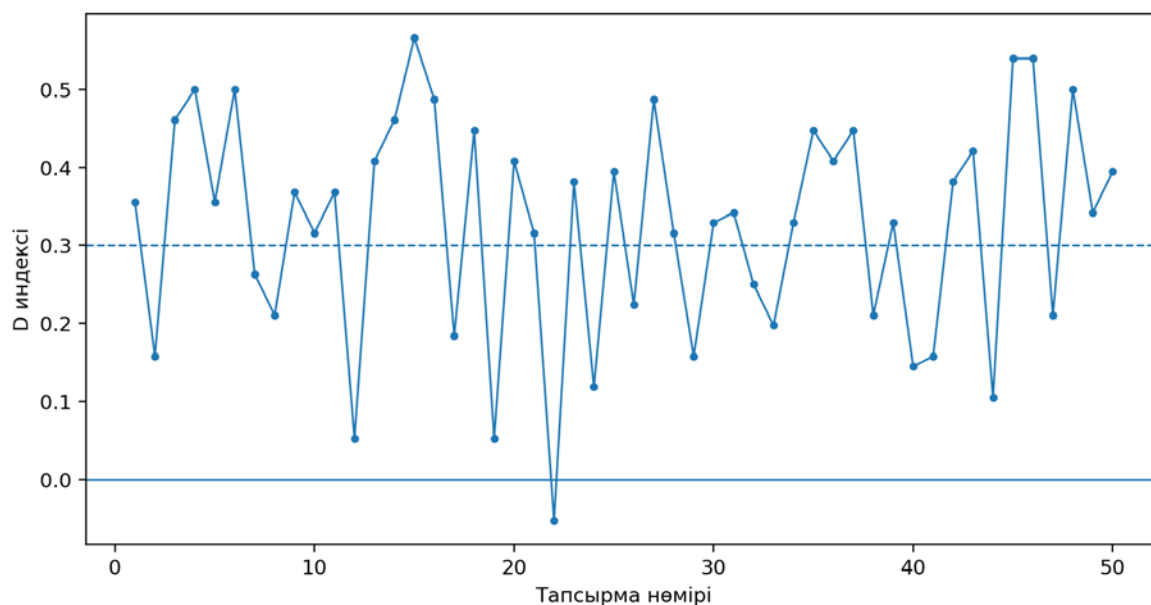
2 сурет - Тапсырмалардың қиындық деңгейі



3 сурет - Тапсырмалар бойынша қиындық индексі

Саралау және тапсырма–тест байланысына тоқталатын болсақ.

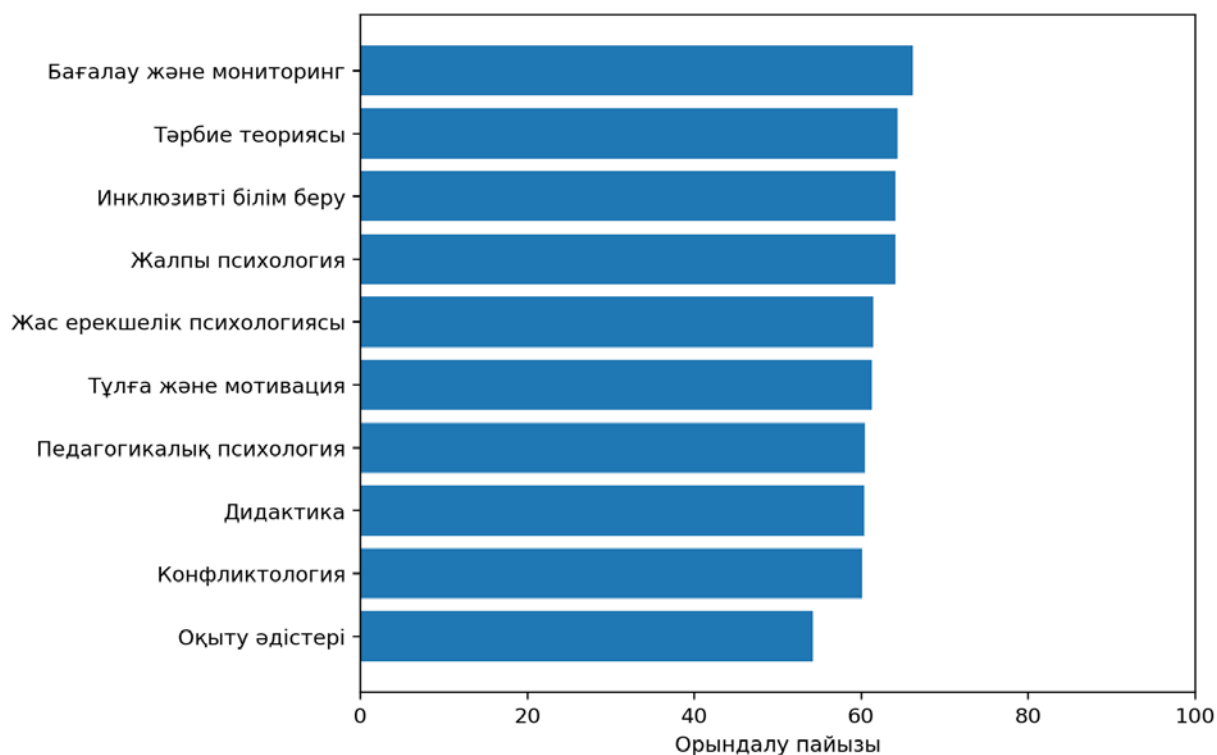
$D \geq 0,30$ деңгейіне 33 тапсырма жетті. $D < 0,20$ аралығында 10 тапсырма, ал теріс саралау көрсеткіші бар 1 тапсырма анықталды. $RIR \geq 0,30$ көрсеткен тапсырмалар саны – 12; $RIR < 0,20$ – 14; теріс RIR – 2. Теріс D немесе RIR анықталған тапсырмалардың дұрыс жауап кілті, тұжырымдамасы және жауап нұсқалары міндетті түрде тексерілуі тиіс.



Кесте 4 - Тақырыптық нәтижелер

Тақырып	Тапсырма саны	Орташа балл	Орындалу
Дидактика	7	4.22	60.4%
Оқыту әдістері	6	3.25	54.2%
Тәрбие теориясы	5	3.22	64.4%
Бағалау және мониторинг	6	3.97	66.2%
Инклюзивті білім беру	5	3.20	64.1%
Жалпы психология	5	3.20	64.1%
Жас ерекшелік психологиясы	5	3.07	61.4%
Педагогикалық психология	5	3.02	60.4%
Тұлға және мотивация	3	1.84	61.3%
Конфликтология	3	1.80	60.1%

Кесте-4 бойынша негізгі байқау. Ең жоғары нәтиже «Бағалау және мониторинг» тақырыбында (66.2%), ең төмен нәтиже «Оқыту әдістері» тақырыбында (54.2%) байқалды. Төмен көрсеткішті бөлімдер біліктілікті арттыру курстарының мазмұнын жоспарлауда басым бағыт ретінде қарастырылуы мүмкін.

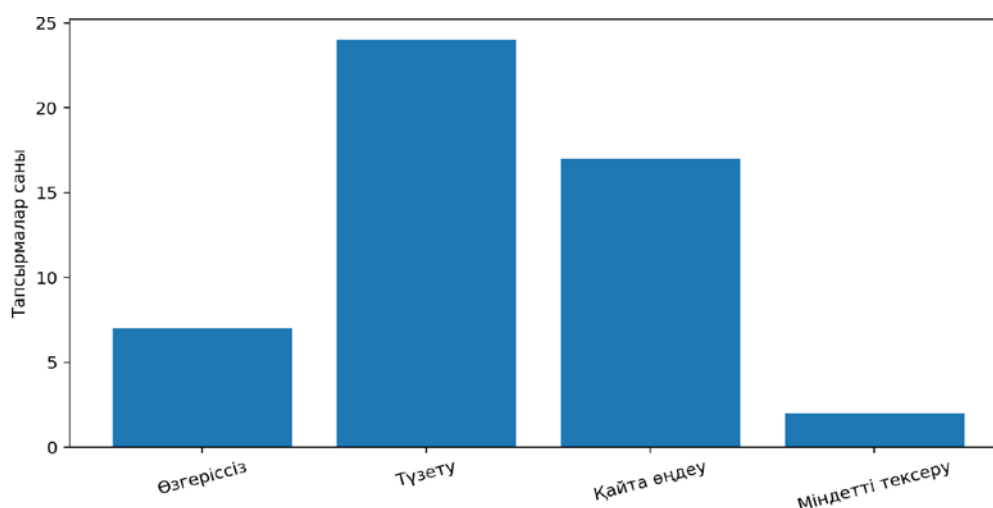


6 сурет – Тақырыптар бойынша меңгеру деңгейі

Кесте 5 - Тапсырмалар бойынша шешімдер

Шешім	Тапсырма саны	Үлесі
Өзгеріссіз	7	14.0%
Түзету	24	48.0%
Қайта өңдеу	17	34.0%
Міндетті тексеру	2	4.0%

Кесте-5 сараптама нәтижесі бойынша тапсырмалардың 82.0%-ы (Түзету мен Қайта өңдеудің қосындысы, барлығы 41 тапсырма) белгілі бір деңгейде қайта қарауды қажет етеді. Бұл жұмыстардың сапасын арттыру үшін қосымша уақыт пен ресурс бөлу керектігін көрсетеді.



7 сурет – Тақырыптар бойынша қорытынды шешімдер

Кесте 6 - Қайта қарауды қажет ететін тапсырмалар

№	Тақырып	P	D	RIR	Тиімсіз дистр.	Шешім
2	Дидактика	0.954	0.158	0.240	3	Қайта өңдеу
3	Дидактика	0.746	0.461	0.351	2	Қайта өңдеу
4	Дидактика	0.643	0.500	0.347	2	Қайта өңдеу
7	Дидактика	0.446	0.263	0.131	0	Қайта өңдеу
8	Оқыту әдістері	0.325	0.211	0.141	1	Қайта өңдеу
12	Оқыту әдістері	0.693	0.053	0.020	2	Кілт пен мазмұнды міндетті тексеру
17	Тәрбие теориясы	0.929	0.184	0.260	2	Қайта өңдеу
19	Бағалау және мониторинг	0.857	0.053	0.027	1	Қайта өңдеу
22	Бағалау және мониторинг	0.782	-0.053	-0.152	3	Кілт пен мазмұнды міндетті тексеру
24	Бағалау және мониторинг	0.925	0.118	0.161	3	Қайта өңдеу
26	Инклюзивті білім беру	0.339	0.224	0.138	1	Қайта өңдеу
29	Инклюзивті білім беру	0.693	0.158	0.093	0	Қайта өңдеу
32	Жалпы психология	0.354	0.250	0.093	1	Қайта өңдеу
33	Жалпы психология	0.921	0.197	0.253	2	Қайта өңдеу
38	Жас ерекшелік психологиясы	0.893	0.211	0.245	2	Қайта өңдеу
40	Педагогикалық психология	0.368	0.145	0.055	0	Қайта өңдеу
41	Педагогикалық психология	0.554	0.158	0.064	1	Қайта өңдеу
44	Педагогикалық психология	0.961	0.105	0.193	3	Қайта өңдеу
47	Тұлға және мотивация	0.407	0.211	0.113	1	Қайта өңдеу

Кестедегі 19 тапсырманың 2-еуі (№12, №22) шұғыл түрде дұрыс жауап кілтін қайта тексеруді, ал қалған 17-сі мазмұнын қайта өңдеуді (жеңіл сұрақтарды күрделендіру, дистракторларды ауыстыру, логикалық екіұштылықты жою) талап етеді.

- Кронбах альфасын кемінде 0,80 деңгейіне көтеру үшін теріс немесе өте төмен RIR көрсеткен тапсырмаларды қайта қарап, пилоттық тексеруді қайталау.

- Өте жеңіл және өте қиын тапсырмалардың үлесін тест спецификациясымен сәйкестендіріп, негізгі бөлікті $P=0,30-0,70$ аралығында сақтау.

- $D<0,20$ тапсырмалардың мәтінін, жауап кілтін және мазмұндық біркәнділігін сараптамалық комиссияда тексеру.

- Екі немесе одан көп тиімсіз дистракторы бар тапсырмалардағы қате жауап нұсқаларын күшейту немесе ауыстыру.

- «Оқыту әдістері» бағыты бойынша педагогтерге арналған мақсатты біліктілікті арттыру модулін қарастыру.

- Нақты деректер алынғаннан кейін нәтижелерді жыныс, өңір, еңбек өтілі және біліктілік санаты бойынша салыстырып, DIF немесе топтық айырмашылықтарды тексеру [13, 14].

Қорытынды және ұсыныстар

Симуляциялық зерттеу тесттің жалпы құрылымы жұмыс істейтінін, бірақ оның сенімділігін және тапсырмалардың бір бөлігінің саралау сапасын жақсарту қажеттігін көрсетті. Психометрикалық сүзгі нәтижелері тапсырма банкін дәлелді түрде жаңартуға, әлсіз тақырыптарды анықтауға және келесі пилоттық тестілеуді мақсатты жоспарлауға мүмкіндік береді.

Психометрикалық талдау техникалық және кәсіптік білім беру жүйесіндегі бағалау сапасының үзілмес кепілі. Педагогика мен психология негіздері бойынша тест тапсырмаларын жүйелі түрде психометриялық сүзгіден өткізу білім беру деңгейін халықаралық стандарттарға жақындатады [15]. Бұл тек педагогтердің білімін әділ бағалап қана қоймай, оқытушылардың кәсіби деңгейін арттыруға және білім беру бағдарламаларын коррекциялауға мүмкіндік беретін тиімді құрал болып табылады.

Әдебиеттер тізімі

1. Шваб К. Четвертая промышленная революция / К. Шваб ; перевод с английского. – Москва : Эксмо, 2016. – 208 с. – URL: <https://eksmo.ru/book/chetvertaya-promyshlennaya-revolutsiya-ITD819325>
2. Аванесов В. С. Форма тестовых заданий : учебное пособие / В. С. Аванесов. – 2-е изд., перераб. и доп. – Москва : Центр тестирования, 2005. – 156 с.
3. Қазақстан Республикасы Оқу-ағарту министрінің 2025 жылғы 25 ақпандағы «Педагогтерді аттестаттаудан өткізу қағидалары мен шарттарын бекіту туралы» № 32 бұйрығы. – ҚР Нормативтік құқықтық актілерінің ақпараттық-құқықтық жүйесі «Әділет», 2025. – URL: <https://adilet.zan.kz/kaz/docs/V2500035751>
4. Cronbach L. J. Coefficient alpha and the internal structure of tests / L. J. Cronbach // Psychometrika. – 1951. – Vol. 16, no. 3. – P. 297-334. – <https://doi.org/10.1007/BF02310555>.
5. Deissinger T. The German system of vocational education and training: Its structural characteristics and current challenges / T. Deissinger // Journal of Education and Work. – 2015. – Vol. 28, no. 3. – P. 225-248.
6. Gassner O. The Quality of Teacher Education in Europe: Performance of National Teaching Standards / O. Gassner, M. Schratz, K. Gruber. – Bucharest : Editura Universităţii din Bucureşti, 2010. – 312 p.
7. Darling-Hammond L. Teacher and Leader Effectiveness in High-Performing Systems: Singapore and Ontario / L. Darling-Hammond, R. Rothman. – Washington, DC : Alliance for Excellent Education, 2011. – 54 p.
8. Crocker L. Introduction to Classical and Modern Test Theory / L. Crocker, J. Algina. – Mason, OH : Cengage Learning, 2008. – 527 p.

9. Haladyna T. M. Developing and Validating Multiple-Choice Test Items / T. M. Haladyna, M. C. Rodriguez. – 4th ed. – New York : Routledge, 2013. – 400 p. – <https://doi.org/10.4324/9780203850381>
10. Kline P. A Handbook of Test Construction (Psychology Revivals): Introduction to Psychometric Design / P. Kline. – London : Routledge, 2015. – 282 p.
11. Field A. Discovering Statistics Using IBM SPSS Statistics / A. Field. – 5th ed. – London : SAGE Publications, 2018. – 1104 p.
12. DeVellis R. F. Scale Development: Theory and Applications / R. F. DeVellis, C. T. Thorpe. – 5th ed. – Thousand Oaks, CA : SAGE Publications, 2021. – 328 p.
13. Ebel R. L. Essentials of Educational Measurement / R. L. Ebel, D. A. Frisbie. – 5th ed. – Englewood Cliffs, NJ : Prentice-Hall, 1991. – 370 p.
14. Birenbaum M. Assessment 2000: Towards a Pluralistic Approach to Assessment / M. Birenbaum // Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge / ed. by M. Birenbaum, F. J. R. C. Dochy. – Dordrecht : Kluwer Academic Publishers, 1996. – P. 3–29. – DOI: https://doi.org/10.1007/978-94-011-0657-3_1
15. Нұрахметов Н. Н. Педагогикалық және психологиялық диагностика негіздері / Н. Н. Нұрахметов, А. А. Реан. – Алматы : Білім, 2018. – 224 б.

Ә.Қ. Құлмағанбет, Д.С. Оспанов, Т.С. Ерсұлтанова

ПСИХОМЕТРИЧЕСКИЙ АНАЛИЗ ТЕСТИРОВАНИЯ В ПЕДАГОГИКЕ И ПСИХОЛОГИИ: ГАРАНТИЯ КАЧЕСТВА ОЦЕНИВАНИЯ В СИСТЕМЕ ТЕХНИЧЕСКОГО И ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

В данном исследовании рассматривается вопрос повышения качества оценки профессиональной компетентности педагогов организаций технического и профессионального образования и обучения (ТиПО) в условиях глобальных вызовов Индустрии 4.0 и цифровизации образования. В контексте реформ, реализуемых на основании приказа Министерства просвещения Республики Казахстан № 32 от 25 февраля 2025 года, обосновывается научно-методическая значимость психометрического анализа (проводимого на макро- и микроуровнях) тестовых инструментов по дисциплинам «Педагогика» и «Психология». В ходе исследования был проведен математико-статистический анализ результатов тестирования 280 участников и 50 тестовых заданий с акцентом на ключевые показатели классической теории тестирования (СТТ): коэффициент альфа Кронбаха (α), валидность, трудность задания (P), индекс дискриминативности (D) и эффективность дистракторов. Был выполнен сравнительный анализ с международными моделями оценивания (Германия, Великобритания, Сингапур, США) и определены направления модернизации отечественной системы оценивания. Психометрический мониторинг продемонстрировал высокую эффективность при оценке качества тестов, а также при выявлении неоднозначных вопросов и нерабочих дистракторов. Результаты исследования позволяют качественно обновлять банк тестовых заданий, определять конкретные профессиональные пробелы у педагогов. Таким образом, психометрический фильтр стал ключевой гарантией объективности оценивания и академической честности в системе ТиПО.

Ключевые слова: ТиПО, психометрический анализ, альфа Кронбаха, индекс трудности, анализ дистракторов, качество оценивания.

A. Kulmaganbet, D. Ospanov, T. Ersultanova

PSYCHOMETRIC ANALYSIS OF TESTING IN PEDAGOGY AND PSYCHOLOGY: ENSURING ASSESSMENT QUALITY IN THE TECHNICAL AND VOCATIONAL EDUCATION SYSTEM

This study examines the issue of improving the quality of professional competence assessment for educators in Technical and Vocational Education (TVE) institutions amidst the global challenges of Industry 4.0 and the digitalization of education. Against the backdrop of reforms implemented pursuant to Order No. 32 of the Ministry of Education of the Republic of Kazakhstan dated February 25, 2025, the study

substantiates the scientific and methodological significance of psychometric analysis-conducted at both macro and micro levels-of assessment tools for the disciplines of Pedagogy and Psychology. The study involved a mathematical and statistical analysis of test results from 280 participants and 50 test items, focusing on key metrics of Classical Test Theory (CTT): Cronbach's alpha (alpha), validity, item difficulty (P), the discrimination index (D), and distractor efficiency. A comparative analysis with international assessment models (Germany, the UK, Singapore, and the USA) was conducted, and directions for modernizing the domestic assessment system were identified. Psychometric monitoring has demonstrated high effectiveness in evaluating test quality and identifying ambiguous questions as well as ineffective distractors. The study results enable the effective updating of the test item bank and the identification of specific professional knowledge gaps among educators. Thus, the psychometric filter has become a key guarantee of assessment objectivity and academic integrity within the TVE system.

Keywords: TVE, psychometric analysis, Cronbach's alpha, difficulty index, distractor analysis, assessment quality.

References

1. Schwab, K. (2016). *Chetvertaya promyshlennaya revolyutsiya* [The fourth industrial revolution] (p. 208). Eksmo. <https://eksmo.ru/book/chetvertaya-promyshlennaya-revolyutsiya-ITD819325>
2. Avanesov, V. S. (2005). *Forma testovykh zadaniy: Uchebnoye posobiye* [Format of test items: A textbook] (2nd ed., p. 156). Tsentr testirovaniya.
3. Ministry of Education of the Republic of Kazakhstan. (2025). *Pedagogterdi attestattdan otkizu qagidalary men sharttaryn bekitu turaly* [On approval of the rules and conditions for conducting certification of educators] (Order No. 32). Adilet Information and Legal System of Regulatory Legal Acts of the Republic of Kazakhstan. <https://adilet.zan.kz/kaz/docs/V2500035751>
4. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
5. Deissinger, T. (2015). The German system of vocational education and training: Its structural characteristics and current challenges. *Journal of Education and Work*, 28(3), 225-248.
6. Gassner, O., Schratz, M., & Gruber, K. (2010). *The quality of teacher education in Europe: Performance of national teaching standards* (p. 312). Editura Universității din București.
7. Darling-Hammond, L., & Rothman, R. (2011). *Teacher and leader effectiveness in high-performing systems: Singapore and Ontario* (p. 54). Alliance for Excellent Education.
8. Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory* (p. 527). Cengage Learning.
9. Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating multiple-choice test items* (4th ed., p. 400). Routledge. <https://doi.org/10.4324/9780203850381>
10. Kline, P. (2015). *A handbook of test construction (Psychology Revivals): Introduction to psychometric design* (p. 282). Routledge.
11. Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed., p. 1104). SAGE Publications.
12. DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed., p. 328). SAGE Publications.
13. Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed., p. 370). Prentice-Hall.
14. Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3–29). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-0657-3_1
15. Nurakhmetov, N. N., & Rean, A. A. (2018). *Pedagogikalyq zhane psikhologiyalyq diagnostika negizderi* [Fundamentals of pedagogical and psychological diagnostics] (p. 224). Bilim.

Авторлар туралы мәлімет:

Құлмағанбет Әсем Қайнолдақызы (автор-корреспондент) – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Техникалық және кәсіптік білім берудегі тест тапсырмасын қалыптастыру ғылыми-әдістемелік зертханасының жетекші сарапшысы, Астана, Қазақстан, e-mail: asemkulmaganbetova@mail.ru

Оспанов Динмухамед Сакенович – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Ғылым мен нарықтық дамыту бөлімінің жетекші ғылыми қызметкері, Астана, Қазақстан, e-mail: dimokospan@gmail.com

Ерсултанова Турсынкуль Серикбаевна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Педагогтердің білімін бағалау тест тапсырмасын қалыптастыру басқармасының басшысы, Астана, Қазақстан, e-mail: tursynkul-07@mail.ru

Сведения об авторах:

Құлмағанбет Әсем Қайнолдақызы (автор-корреспондент) - РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий эксперт научно-методической лаборатории по формированию тестовых заданий для оценки квалификации педагогов технического и профессионального образования, Астана, Казахстан, e-mail: asemkulmaganbetova@mail.ru

Оспанов Динмухамед Сакенович – РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий научный сотрудник отдела науки и рыночного развития, Астана, Казахстан, e-mail: dimokospan@gmail.com

Ерсултанова Турсынкуль Серикбаевна - РГП на ПХВ «Национальный центр тестирования» МНВО РК, руководитель управления формирования тестовых заданий оценки знаний педагогов, Астана, Казахстан, e-mail: tursynkul-07@mail.ru

Information about authors:

Kulmaganbet Assem (corresponding author) - Master's, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education Republic of Kazakhstan, Lead Expert at the Scientific and Methodological Laboratory for the Development of Test Items to Assess the Qualifications of Technical and Vocational Education Teachers, Astana, Kazakhstan, e-mail: asemkulmaganbetova@mail.ru

Ospanov Dinmukhamed – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education Republic of Kazakhstan, Leading Researcher of the Science and Market Development Department, Astana, Kazakhstan, e-mail: dimokospan@gmail.com

Ersultanova Tursynkul – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education Republic of Kazakhstan, Head of the Department for Developing Test Items for Assessing Teacher Knowledge, Astana, Kazakhstan, e-mail: tursynkul-07@mail.ru

N. Issina^{1*}, A. Yerzhumanova², G. Baideldinova³

^{1,2,3}Abylka Saginov Karaganda Technical University, Karaganda, Kazakhstan

*e-mail: nazerkeissina01@gmail.com

¹ORCID 0000-0001-5412-1527, ²ORCID 0009-0008-7910-0712,

³ORCID 0000-0002-2905-1130

GENERATION Z STUDENTS' PREFERENCES IN LEARNING EFL IN KAZAKHSTAN: AN ONLINE SURVEY STUDY

This study analyses the preferences of Generation Z students (first-year students, aged 17-18) regarding the study of English as a foreign language at universities in Kazakhstan. The aim of this study is to identify the opinions of first-year students (aged 17-18), representatives of Generation Z, regarding their preferred methods of learning English as a foreign language, as well as to determine the factors that influence their interest and engagement in the educational process. Empirical basis: an online survey of 93 students, analysed using descriptive statistics and qualitative analysis of open-ended responses. Particular attention is paid to interest and academic engagement, the role of social media, preferences for interactive/visual formats, attitudes towards AI tools, as well as characteristics of learning behaviour – from time management and concentration to expectations regarding feedback. The results demonstrate enthusiasm for visual and interactive methods, social media and digital tools: students prefer short tasks and rapid feedback, which fits perfectly with their dynamic learning style. Attitudes towards AI are mixed – they recognise its benefits but fear a loss of autonomy and critical thinking. The conclusions emphasise that it is time to adapt English teaching to the digital habits of Zoomers. Practical value lies in recommendations for curricula and the integration of technology into language education.

Keywords: generation Z, language learning, students' preferences, Kazakhstani universities, social media in language learning, AI in education, language education in Kazakhstan.

Introduction

In the current context of digital transformation in education, the student body has changed significantly. At present, the majority of students in higher education institutions are Generation Z (born approximately between the mid-1990s and the early 2010s), while most teachers belong to the millennial generation. This intergenerational difference creates certain pedagogical challenges, as the values, cognitive characteristics and ways of interacting with information of these generations differ significantly.

Generation Z is characterised as the first generation of 'digital natives' who grew up with constant access to the internet, mobile devices and social networks. Representatives of this generation are characterised by high-speed information processing, a preference for visual and multimodal content, a focus on short formats for presenting material, a need for immediate feedback, and a habit of interactive communication. At the same time, researchers note a reduced attention span, a tendency towards fragmented perception of information, and high sensitivity to the pace of learning.

In contrast, millennials – a generation that grew up during the rise of digital technology – are more often oriented towards text-based learning formats, longer educational blocks, and traditional models of academic interaction. Despite their high digital adaptability, millennials' teaching practices often rely on methodological approaches formed in the pre-digital era. This can lead to a mismatch between the expectations of teachers and students, a decrease in the latter's motivation, and a reduction in their level of involvement in the learning process.

This problem is particularly relevant in the context of teaching English as a foreign language at a technical university. For first-year students, English language proficiency is not only an academic discipline, but also a tool for professional mobility, access to international sources of information,

and participation in global communication. In a rapidly changing digital environment, there is a need to adapt foreign language teaching methods to the characteristics of Generation Z.

Despite the existence of studies devoted to the educational characteristics of 'Zoomers,' the question of which methodological formats students themselves consider most effective in the process of learning a foreign language remains insufficiently studied, especially in the regional context of Kazakhstan and in the environment of a technical university.

The aim of this study is to identify the opinions of first-year students (aged 17-18), representatives of Generation Z, regarding their preferred methods of learning English as a foreign language, as well as to determine the factors that influence their interest and engagement in the educational process.

The study is based on the following hypothesis: the use of social networks, interactive and visual teaching methods, as well as AI tools, taking into account students' preferences in terms of time, concentration and speed of feedback, has a positive effect on their interest and involvement in the process of learning English.

To test the hypothesis, an empirical study will be conducted in the form of an online survey among first-year students at Abylkas Saginov Karaganda Technical University who are studying English at the Elementary and Pre-Intermediate levels, followed by statistical analysis of the data obtained.

Thus, the study aims to refine pedagogical strategies for teaching English, taking into account the intergenerational characteristics and digital preferences of the modern student audience, which can contribute to improving the effectiveness of the educational process in higher education.

In accordance with the stated relevance of the research, the formulated objective and the proposed hypothesis, it is necessary to analyse existing scientific approaches and empirical data on this issue. A review of theoretical positions and the results of contemporary research will allow us to determine the degree to which the topic has been developed, identify existing scientific positions, and justify the conceptual basis of this study. In this regard, the following is an overview of the scientific literature devoted to key aspects of the issue under study.

The research presented by Yohanes Agung Wahyu Nugroho, Bonita Eka Dama Yanti and Felik Haryanto explore the multifaceted nature of Generation Z, analysing their unique characteristics in the areas of education, work and mental well-being. The authors view 'digital natives' as a category whose worldview is shaped by constant access to technology, a pragmatic attitude towards finance, and a desire for inclusivity. As members of this generation grew up in an era of rapid technological progress, their educational expectations are inextricably linked to technology [1].

According to Hülya Düzenli, Generation Z (born between the mid-1990s and early 2010s) are the first 'digital natives' for whom the internet and mobile technologies are a natural habitat. Their main educational characteristics include:

- Short attention spans and a habit of obtaining information quickly.
- A preference for visual forms of learning over text-based ones.
- High technical competence, which often exceeds the skills of their teachers.
- A need for instant feedback, answers, and rewards.

Hülya Düzenli classifies educational recommendations into five main categories: speed, motivation, content, learning environment, and active learning.

For Generation Z, the pace of interaction is critical. Instant feedback is the most frequently repeated recommendation in education. Students expect faster guidance and responses from instructors than previous generations. Teaching Gen Z is inextricably linked to their social values: they find environments that support inclusive and multicultural education more useful. It is necessary to move from rigid hierarchical models to a more democratic, technology-driven and flexible environment [2].

The study by Goran Ćurić, Nikola Papac and Božana Prusina revealed that Generation Z as 'digital natives' and 'the one-click generation,' whose knowledge comes almost entirely from digital sources. Their attention is selective and short-lived; they tend to quickly filter out information and lose interest if the content does not seem relevant to them.

According to the article, the education system and business environment must adapt to the 'hyper-networked' nature of this generation's lives. The key to their successful learning lies in combining technology (digital environment, gamification) with a deeply human approach (empathic mentoring, personalisation) [3].

Elizabeth Frissell's article 'The Dangers of Social Media and the Stoic Solution' (2020) is a philosophical and critical analysis of the influence of influencer culture on Generation Z. The author argues that the 'media monopolies' of the 20th century have been replaced by a new power – opinion leaders who shape the values and behavioural patterns of young people through digital platforms.

The article examines education in the context of life values and proposes 'stoic pedagogy' as a method of self-help:

- Conflict of values: The author notes that the desire to meet the standards of influencers (buying branded items, chasing external glamour) distracts resources. Money and time that could be spent on education and career development are spent on impulsive consumption for the sake of social approval.

- Rejection of traditional sources of knowledge: Generation Z rejects old forms of media, considering them 'depressing,' and is completely transitioning to a digital environment where information is presented in an entertaining way [4].

The research by Cain C.C., Morgan Bryant, etc. is devoted to the development of a conceptual model Academic Technology Framework, which aims to bridge the gap between the teaching methods of professors and the educational expectations of Generation Z students.

The authors define zoomers as 'true digital natives' who do not remember a world without smartphones and instant access to information. Zoomers are more likely to attend college and less likely to drop out than millennials. For them, Google is the primary source of information, and the professor is only secondary. This fundamentally changes the traditional dynamics of learning.

The article argues that teaching Generation Z requires a shift away from the traditional 'lecture' format in favour of a technology-rich, flexible, and practice-oriented environment where technology is not an add-on to the process but an integral part of it [5].

The research paper by Philipp and Sulkovsky is devoted to the transformation of the professional environment under the influence of artificial intelligence and the peculiarities of Generation Z's adaptation to these changes.

The authors argue that learning for Zoomers should balance high technology with human involvement (high-tech yet high-touch). An important task in education is to preserve the individuality of the learner. Programmes should encourage the use of AI as a draft that requires mandatory editing to express one's own 'voice' and authorship [6].

The article by Carissa Ullman is a practical guide for higher education institutions on adapting to the needs of Generation Z. Unlike millennials, Gen Zers are defined by the author not simply as 'digital nomads' but as 'digital innates' whose perception of the world and education has been shaped by global instability and economic decline.

According to sources, higher education is currently 'behind the times'. To successfully educate Generation Z, institutions must move to a partnership model of learning and link each academic task to the development of real-life skills, as well as ensure a balance between modern digital portals and in-depth personal mentoring.

Zoomers are 'difficult customers' for universities, but their desire for self-fulfilment and responsibility makes them some of the most promising students, if the educational institution is willing to adapt to their pace and values [7].

Brenda van Wijk examines the technological preferences of a new generation of students and analyses how well higher education is prepared to meet their needs.

The key concept of the article is "phygital" – the concept of using technology to bridge the gap between the digital and physical worlds to create a unique interactive experience. Students of this generation do not see a clear difference between physical and virtual space. They prefer to navigate

all aspects of their lives, including study and research, exclusively using smartphones and mobile devices.

According to the author, academic staff and support services at higher education institutions often feel overwhelmed and unprepared for the demands of “figital students”. To bridge this gap, it is necessary to rethink curricula, integrate mobile technologies as a primary learning tool, and transition to a student-centred model where technology facilitates deep learning and critical thinking [8].

Dr. Hernando Jr Bernal addresses the need to transform the education system in the face of rapid technological progress and generational change among students. According to the source, Generation Z grew up at a time when the internet and technological advances were just beginning to grow rapidly. Unlike the next generation (Alpha), for whom technology is an innate part of life, Generation Z caught the moment when these technologies were new, and teachers taught them how to use them [9].

For effective teaching of zoomers, Bernal suggests moving away from rote memorisation of facts and implementing three key tactics: facilitated learning, learning integration, and soft skills training: In a world of instant answers, it is critical to teach perseverance, the ability to manage impulsivity, and an understanding of ‘why’ we learn, not just ‘what.’

Analysis of the article shows that teaching Generation Z requires an evolution in teaching style: a shift from content delivery to the development of higher-order cognitive skills. It is important to use students' innate technological knowledge to create curricula based on research and collaboration [10].

The research work by Hala Shaker Hammad is a comprehensive study of how Generation Z perceives the educational process. The study is based on a survey of 139 students and 29 teachers at a private university in Egypt.

The article highlights several fundamental requirements that Zoomers have for the learning process. About 50% of students consider multimedia elements to be critical for understanding the material. 59% of Zoomers prefer to learn through YouTube videos. Presenting information in small segments (chunks) rather than long lectures is recognised as an effective method. 59% of students expect detailed and constructive feedback that will help them understand their strengths and areas for growth [11].

The main conclusion of the article is that higher education should move away from traditional lecture methods towards a student-centred, flexible and technology-rich model. Universities should prepare zoomers for professions that do not yet exist, focusing on the development of critical thinking, curiosity and practical skills [12].

Szabó Cs. M., Bartal O. and Nagy B. define modern students as ‘digital natives’ whose values and problem-solving methods differ fundamentally from those of previous generations. At the same time, despite their natural proficiency with gadgets, they may have difficulty with in-depth understanding of texts, online ethics, and filtering fake news.

The article emphasises that although Generation Z is technologically advanced, their educational needs remain deeply practical. What matters to them is not technology itself, but how it helps them gain real skills for the labour market. At the same time, the digital environment creates risks for deep learning, which requires educators to implement strategies that support concentration and critical analysis [13].

An analysis of scientific sources shows that contemporary pedagogical literature provides a fairly detailed description of the cognitive, motivational, and digital characteristics of Generation Z, as well as justifying the need to adapt educational practices to their learning preferences. Researchers emphasise the importance of high interactivity, visualisation of educational content, the use of digital platforms, gamification and rapid feedback in the learning process. At the same time, most existing studies are general in nature and focus primarily on foreign educational contexts.

At the same time, there remains insufficient research on how Generation Z students in Kazakhstan perceive the process of learning English as a foreign language, what formats and platforms they consider most effective, and what role they believe digital technologies and social

networks play in language learning. There is a particularly limited number of empirical studies based on direct surveys of first-year students aged 17-18 who are in the early stages of higher and post-secondary education.

In this regard, the present study aims to obtain empirical data on the preferences of Generation Z students in learning English as a foreign language through an online survey. The following section describes the research methods, including sample characteristics, instruments, data collection procedures, and methods of analysis.

Materials and methods

This study is combined in nature and includes theoretical and empirical methods of cognition. The study was conducted within the framework of a quantitative descriptive design with elements of qualitative analysis and is based on an online survey of students.

The following theoretical methods were used:

- Analysis and synthesis of scientific literature on the problems of teaching foreign languages to Generation Z;
- Comparative analysis of studies devoted to the characteristics of Generation Z in the educational process;
- Generalisation of theoretical propositions related to digital learning, motivation and preferences of learners.

The theoretical analysis made it possible to identify the key characteristics of Generation Z as learners, as well as to reveal the insufficient study of their learning preferences in the context of teaching English as a foreign language in Kazakhstan.

Research participants

The empirical study was conducted among first-year students at Abylkas Saginov Karaganda Technical University who are studying English as a foreign language as part of the compulsory curriculum.

The study involved students aged 17-18, belonging to Generation Z, with Elementary, Pre-Intermediate and Intermediate English language proficiency (A1-B1 according to CEFR).

Participation in the survey was voluntary and anonymous. All respondents were informed about the purpose of the study before the survey began.

The main tool for collecting empirical data was an online questionnaire developed by the authors of the study. The questionnaire included:

- Closed questions using a five-point Likert scale;
- Multiple-choice questions;
- Open-ended questions allowing students to express their own opinions.

The aim of the survey was to identify the preferences of first-year students, representatives of Generation Z, in learning English as a foreign language. In particular, the study focused on analysing students' attitudes towards the use of social networks, interactive and visual teaching methods, artificial intelligence tools, as well as studying the characteristics of their learning behaviour related to time management, concentration and expectations regarding the speed of feedback in the learning process.

The survey was conducted online, which ensured its accessibility and allowed all research participants to be included.

The data obtained was processed using quantitative and qualitative analysis methods. Descriptive statistics methods were used to analyse the survey results, including:

- calculation of frequencies and percentages;
- calculation of mean values.

These methods made it possible to identify general trends in student motivation and learning preferences.

Students' open-ended responses were analysed using thematic analysis, which allowed us to identify the most frequently recurring semantic categories and supplement the quantitative data with qualitative interpretations.

The reliability of the study was ensured by:

- A standardised questionnaire form;
- Unified conditions for conducting the online survey;
- The use of proven assessment scales.

The validity of the study was ensured by the correspondence of the questionnaire content to the goals and objectives of the study, as well as by reliance on theoretical provisions identified in the course of the literature review.

Results and discussion

1. The Socio-Demographic Data

93 first-year students participated in the study. Most participants were 18-year-old students (50.5%), slightly fewer were 17-year-olds (49.5%). Men predominated in terms of gender (65.6%). Most students have been studying English for less than 5 years (55.9%), while 29% have 5-8 years of experience and 15.1% have more than 8 years. The level of English proficiency among most participants is Elementary (68%) and Pre-Intermediate (19.4%), with only a small proportion at Intermediate (10.8%) and Upper-Intermediate (1.1%).

2. Students' Interest and Academic Engagement in Learning English

Table 1 – Students' Interest and Academic Engagement in Learning English

Question	1	2	3	4	5	4+5 (Agreement)
I am interested in learning English	4,3%	4,3%	21,5%	24,7%	45,2%	69,9%
I actively participate in tasks during English classes	2,2%	3,2%	20,4%	40,9%	33,3%	74,2%
I enjoy doing additional English assignments	6,5%	15,1%	37,6%	21,5%	19,4%	40,9%
Modern lesson formats make learning English more engaging	3,2%	6,5%	16,1%	22,6%	51,6%	74,2%

Table 1 shows the analysis of the results:

- Most students (≈ 70 -74%) show a high level of interest in learning English and actively participate in class assignments (questions 1, 2 and 4).
- Satisfaction with modern classroom formats is also high, confirming the positive impact of interactive and visual techniques.
- Interest in completing additional assignments outside of class is less pronounced: only 40.9% of students indicated agreement, which may indicate a need for motivational strategies for work outside of class.

These results confirm that Generation Z students are actively involved and interested in the learning process, especially when modern teaching formats are used.

3. Use of Social Networks in English Language Learning

To assess the impact of social networks on the process of learning English, students responded using a Likert scale (1 – strongly disagree, 5 – strongly agree). The distribution of responses for each question is presented in Table 2.

Table 2 – Use of Social Media in English Language Learning

Question	1	2	3	4	5	4+5 (Agreement)
I learn new English words or grammar structures through content on social media (Instagram, TikTok, etc.)	5,4%	14%	29%	21,5%	30,1%	51,6%
Social media is a source of English language materials for me	4,3%	7,5%	33,3%	28%	26,9%	54,9%
I would like some of my homework or class assignments to be related to social media content	2,2%	8,6%	32,3%	37,6%	19,4%	57%
Using social media in learning English seems reasonable to me	2,2%	9,7%	32,3%	24,7%	31,2%	55,9%
I prefer learning English without using social media	16,1%	20,4%	32,3%	18,3%	12,9%	31,2%

Table 2 shows the analysis of the results:

- More than half of students ($\approx 52\text{-}57\%$) actively use social media to learn English or would like some of their assignments to be related to social media content (questions 1-3).
- Students generally have a positive attitude towards the idea of using social media in education (question 4 – 55.9% agreement).
- Only a small percentage ($\approx 31\%$) prefer to learn English without using social media, which confirms the importance of digital content for Generation Z.

The results show that integrating social media into the learning process can increase student interest and engagement, especially if some of the assignments or content are related to social media.

Results of an open-ended question about the use of social media in learning English

Students answered the question: ‘How would you like social media to be integrated into English lessons? Please give an example.’ Analysis of the responses revealed several main areas of social media use, which can be divided into categories.

1. Use of video content and films

Most respondents noted that watching short videos or films in English helps them understand the material more quickly and increases their interest in learning the language. Examples include: ‘Watching films helps you understand grammar and vocabulary’ and ‘maybe watch the film...’.

2. Interactive tasks and games

Some students suggested integrating games and quizzes through social networks or online platforms. Examples of responses: ‘Quiz games’, ‘games’, ‘Kahoot, etc.’

3. Creating content in English via social media

Students also suggested creating content in English themselves, for example, writing posts on Instagram or shooting short videos on TikTok. Examples: ‘Correspondence via Instagram’, ‘publish some stories in English’, ‘students can make short videos on TikTok or write posts on Instagram in English’.

4. Communication and practice through social media/chat

Some responses mentioned using social media for communication and real-life practice of English. Examples: ‘Teachers create group chats and ask students to write in English,’ ‘I communicate with people from other countries via the internet.’

5. Use of translators and apps

A separate group of students noted that it is convenient to use technological resources such as translators and language learning apps to learn words and phrases. Examples: ‘Google Translate,’ ‘use apps to learn English, for example Duolingo’.

6. Negative or sceptical responses

Some students expressed scepticism or a negative attitude towards the use of social media in learning. Examples: 'No, I don't think that's a good idea because everyone needs a different approach', 'I'm not very interested in using social media'.

Overall, most students demonstrated a positive attitude towards the use of social media in learning English, especially in the form of videos, interactive tasks and creating their own content. The responses show that the integration of modern digital tools can help increase student interest and engagement in the learning process.

4. Students' Preferences for Interactive Methods in English Language Learning

Table 3 – Students' Preferences for Interactive Methods in English Language Learning

Question	1	2	3	4	5	4+5 (Agreement)
I enjoy tasks that require active participation (discussions, games, group work)	3,2%	6,5%	26,9%	26,9%	36,6%	63,5%
I participate more actively in class when doing interactive tasks	1,1%	12,9%	37,6%	28%	20,4%	48,4%
The format of the task affects how easily I understand new material	2,2%	10,8%	24,7%	34,4%	28%	62,4%
I prefer traditional tasks (reading, translation, textbook exercises) over interactive formats	5,4%	11,8%	34,4%	30,1%	18,3%	48,4%

Table 3 shows the analysis of the results. Respondents have a positive attitude towards interactive teaching methods. Most students (63.5%) agreed with the statement that they like tasks that require active participation (discussions, games, group work). When completing interactive assignments, almost half of the respondents (48.4%) noted that they participate more actively in class.

The format of the assignment has an impact on the assimilation of new material: 62.4% of participants noted that interactive elements help them to perceive the learning material more easily. At the same time, traditional tasks (reading, translation, exercises from the textbook) are less popular – 48.4% of students agreed with their advantage, which shows a slight superiority of interactive methods (51.6%) over traditional formats.

Overall, the results indicate high student engagement and a positive perception of interactive teaching methods in English.

The results of an open-ended question.

Responding to the open-ended question, 'What interactive tasks do you like most in English classes?', students indicated a variety of preferences that can be divided into several categories.

1. Role-playing and speaking:

Many respondents noted that they like activities related to real-life communication and speaking practice. Examples include role-playing in a shop, airport or doctor's office, as well as dialogues between students ('One student is the salesperson, the other is the customer').

2. Group work and discussions:

Respondents also highlighted pair or group work, including discussions, debates, and collaborative exercises. These formats allow students to express their thoughts in English and interact with their classmates, which increases their engagement in the learning process.

3. Game formats and quizzes:

Many students mentioned the use of online quizzes and educational games such as Kahoot, Quizziz, and True/False.

4. Visual and multimedia elements:

Some respondents noted that they like video lessons, discussions of short videos, films, and multimedia assignments.

5. Writing and creative tasks:

Some students indicated that they like tasks that involve writing short texts, such as ‘My dream weekend’ or ‘My favourite hobby,’ which are then read aloud and discussed in class.

Thus, the open-ended responses show that students highly value interactive teaching methods, especially role-playing, group discussions, games and multimedia tasks, as they promote practical use of the language and increase motivation.

5. Students’ Preferences for Visual Learning Methods

Table 4 – Students’ Preferences for Visual Learning Methods

Question	1	2	3	4	5	4+5 (Agreement)
Visual materials (videos, images, infographics, presentations) help me remember information better	2,2%	5,4%	30,1%	30,1%	32,3%	62,4%
Using visual tools helps me focus better in class	2,2%	6,5%	37,6%	24,7%	29%	53,7%
Video and visual content make learning English easier to understand	2,2%	7,5%	25,8%	29%	35,5%	64,5%
I find it easier to understand new material when it is accompanied by visual elements	0%	6,5%	34,4%	23,7%	35,5%	59,2%

Table 4 shows the analysis of the results. Students noted that visual materials such as videos, images, infographics, and presentations contribute to better retention of information. Thus, 62.4% of respondents agreed with this statement (ratings 4 and 5). The use of visual aids also helps maintain concentration in class (53.7% agree), facilitates understanding of the material (64.5% agree), and makes it easier to perceive new material when accompanied by visual elements (59.2% agree). These data show that visual teaching methods are an effective tool for improving the perception and retention of information in English lessons.

The results of an open question “Which visual materials or formats in English classes do you like the most?:

Most students prefer visual materials such as videos and presentations in English classes. Short video lessons, videos with dialogues or interesting plots, and presentations that help them better perceive and understand the material were frequently mentioned in the responses. Some respondents noted infographics as a useful format for visual memorisation of information.

Examples of student comments:

- ‘Video lessons help me understand the material faster and make classes more interesting.’
- ‘Presentations and visual explanations from the teacher make lessons lively and understandable.’
- ‘Infographics and videos help me remember new vocabulary and grammar better.’
- ‘Videos with short dialogues improve listening comprehension and pronunciation.’

Thus, visual materials (videos, presentations, infographics) are the most preferred among students and contribute to increasing interest and effectiveness in learning English.

6. Use of Artificial Intelligence in English Language Learning

Table 5 – Use of Artificial Intelligence in English Language Learning

Question	1	2	3	4	5	4+5 (Agreement)
Using AI (e.g., ChatGPT, online translators, language learning apps) increases my interest in learning English	5,4%	11,8%	32,3%	17,2%	33,3%	50,5%
AI helps me better understand English grammar and vocabulary	1,1%	6,5%	44,1%	22,6%	25,8%	48,4%
With the help of AI, it is easier for me to complete assignments independently	4,3%	3,2%	34,4%	29%	29%	58%
I often use AI resources to practice speaking skills, pronunciation, or other aspects of English	5,4%	10,8%	37,6%	19,4%	26,9%	46,3%

Table 5 shows the analysis of the results. In the section devoted to the use of artificial intelligence in learning English, respondents assessed the extent to which AI affects their interest, understanding of grammar and vocabulary, independent completion of tasks, and practice of speaking skills.

The results show that the use of AI generates considerable interest among students: 50.5% of respondents agreed with the statement that AI increases their motivation to learn English. At the same time, almost half of the students (48.4%) noted that AI helps them to better understand grammar and vocabulary. More than half of the survey participants (58%) believe that AI makes it easier to complete tasks independently.

When it comes to practising speaking and pronunciation skills, 46.3% of students use AI resources to improve these aspects of the language. At the same time, a small number of respondents expressed doubts or low interest in using AI, which may be due to individual preferences in learning methods.

Overall, the data indicates a positive perception of AI as a tool that promotes interest in learning English and increases independence in the learning process.

The results of an open-ended question

The analysis of an open-ended question “Do you think that using AI in any way hinders your independence or the development of critical thinking? Please explain” is shown below.

Respondents expressed mixed opinions about the impact of AI use on their independence and critical thinking skills. Some students (about 25-30%) noted that excessive use of AI can create dependency and reduce the ability to think independently. They explained that if tasks are completed entirely with the help of AI, the ability to formulate ideas independently, make decisions and develop creative thinking is reduced. For example, one student noted: ‘Yes, I think it is harmful because we rely on it and stop learning independently’. Another added: ‘If we only perform a task using AI, our ability to form our own thoughts, make decisions, and work creatively may be diminished.’

Some students did not see any negative effects, pointing out that AI makes it easier. Other respondents emphasised the positive impact of AI when used as a support tool. They noted that AI helps them understand material more quickly, improve their grammar and vocabulary, and develop their writing and speaking skills, without hindering their independence when used consciously. As one student noted: “No, it doesn’t interfere; AI complements my ideas.” Another added: “I think AI can help in learning English, but you shouldn’t overdo it.”

Easier to complete complex tasks, but does not replace their own thinking. For example, one respondent wrote: “AI is just a tool. It does not interfere with independence if used consciously”. Several participants also emphasised that the effectiveness of AI depends on how it is used and, on the teacher’s, pedagogical methods.

Overall, the results show that students perceive AI as a useful tool that can both support and potentially limit independent learning and the development of critical thinking if used excessively or incorrectly.

7. Time, Attention, and Speed of Feedback in Learning

Table 6 – Time, Attention, and Speed of Feedback in Learning

Question	1	2	3	4	5	4+5 (Agreement)
I regularly spend time studying English outside of class	9,7%	22,6%	32,3%	18,3%	17,2%	35,5%
I often postpone English assignments if they seem too extensive or demanding	8,6%	11,8%	37,6%	19,4%	22,6%	42%
I find it difficult to stay focused on one task for a long period of time	4,3%	12,9%	32,4%	23,7%	24,7%	48,4%
I usually expect quick feedback from the teacher on completed assignments	10,8%	12,9%	38,7%	17,2%	20,4%	37,6%
I prefer completing several short tasks instead of one large assignment	8,6%	3,2%	34,4%	23,7%	30,1%	53,8%

Table 6 shows the analysis of the results. In the section devoted to time, attention and speed of feedback when learning English, respondents demonstrated a variety of habits and preferences. Only about 35% of students regularly devote time to studying English independently outside of class, while the majority of participants either study irregularly or occupy a middle position on this scale. This indicates that the proportion of those who agree with regular study is relatively low.

When doing homework, only 42% of respondents agree that they often put it off if the assignments seem too voluminous or difficult. At the same time, almost half of the students (48%) noted that they have difficulty concentrating on one task for a long time. These figures show that fewer respondents agree with discipline and concentration than those who choose average or neutral positions.

In terms of feedback, 38% of participants usually expect a quick response from the teacher on completed assignments, and more than half (54%) prefer to complete several short assignments instead of one large one. This demonstrates that there are fewer active respondents who agree with such preferences than students who take neutral positions, but their share is still significant.

Overall, the results show that there are fewer students with clear habits and preferences than those who take a neutral position. Nevertheless, most respondents value flexibility in the organisation of their studies and timely feedback, which contributes to improving the effectiveness of learning and maintaining attention during classes.

8. Time Preferences in English Language Learning

Table 7 – Time, Attention, and Speed of Feedback in English Language Learning

Question	Answer option	%
How much time per day are you willing to spend studying English outside of class?	Less than 15 minutes	11,8%
	15-30 minutes	28%
	30-45 minutes	29%
	45-60 minutes	19,4%
	More than 60 minutes	11,8%
Which lesson duration format is most comfortable for you?	10-15 minutes	5,4%
	20-30 minutes	17,2%
	40-45 minutes	38,7%
	More than 45 minutes	38,7%

Table 7 shows the analysis of the results. The results concerning students' time preferences when studying English show a moderate willingness among learners to devote time to independent work outside of class. The largest proportion of respondents (29%) indicated that they were willing to spend 30-45 minutes a day studying English. A similar number of students (28%) said they could devote 15-30 minutes to this. A smaller proportion of respondents are willing to study for 45-60 minutes (19.4%), while 11.8% of students chose less than 15 minutes, and the same proportion (11.8%) indicated that they are willing to devote more than 60 minutes a day. Overall, the results show that most students are willing to devote approximately 15 to 45 minutes daily to independent English language study.

Students' preferences regarding the duration of a lesson also show a tendency towards medium or longer lessons. Most respondents indicated that the most comfortable format for them is lessons lasting 40-45 minutes (38.7%) or more than 45 minutes (38.7%). Shorter formats were less popular: 17.2% of students chose 20-30 minutes, and only 5.4% chose 10-15 minutes. These data show that most students prefer standard or longer classes, which allow them to delve deeper into the material and complete various types of assignments.

A summary analysis of the data reveals several consistent trends.

Firstly, students have a positive attitude towards interactive and visual teaching methods, as the proportion of those who agree with most of the statements in these blocks exceeds 60%. This indicates that active forms of work, as well as the use of videos, images and presentations, are perceived by students as effective means of assimilating learning material.

Secondly, the results demonstrate a moderately positive attitude towards the use of social networks and artificial intelligence in English language teaching. More than half of the students consider social networks to be a useful source of educational content and support the idea of integrating such resources into educational tasks. Similarly, about half of the respondents note that the use of AI increases interest in learning the language and helps them complete assignments.

At the same time, the data related to learning time and concentration show a lower level of agreement. This indicates that students are not always willing to devote significant time to independent work outside of class and may have difficulty maintaining concentration for long periods of time. At the same time, more than half of respondents prefer several short assignments to one large one, which indicates the importance of breaking down learning activities into small steps.

Conclusion

The results obtained allow us to conclude that the hypothesis put forward has been partially confirmed. The hypothesis suggested that the use of social networks, interactive and visual teaching methods, as well as artificial intelligence tools, taking into account students' preferences regarding time, concentration and speed of feedback, contributes to increasing students' interest and involvement in the process of learning English.

Data analysis showed that interactive and visual teaching methods have the most pronounced positive effect on student interest and engagement. Most respondents noted that tasks requiring active participation (discussions, games, group work) make the learning process more attractive, and the use of videos, presentations and other visual materials contributes to a better understanding and memorisation of the learning material.

The use of social media also received a generally positive assessment from students. More than half of the respondents consider social media to be a source of materials for learning English and support the inclusion of such content in educational assignments. This demonstrates the potential for integrating digital platforms into the educational process.

Students' attitudes towards the use of artificial intelligence tools proved to be more ambiguous. On the one hand, a significant proportion of respondents noted that AI helps them to better understand the course material and makes it easier to complete assignments. On the other hand, in their open responses, students pointed to the possible risks of excessive dependence on AI and a decline in independence if it is used in an uncontrolled manner.

In addition, the results of the study showed the importance of taking into account students' cognitive and organizational preferences. A relatively small proportion of students regularly study English outside of class, and many respondents note difficulties with sustained concentration. At the same time, most students prefer to complete several short assignments rather than one large one, which indicates the need to adapt the structure of learning tasks and the format of feedback.

Thus, the results of the study confirm that the integration of modern digital tools and active teaching methods can help increase student interest and engagement, but their effectiveness largely depends on pedagogically sound and balanced use.

References

1. Yohanes Agung Wahyu Nugroho, Bonita Eka Dama Yanti, Felik Haryanto. A Systematic Literature Review on Workplace Expectations and Behavioral Characteristics of Generation Z EmployeesPsikostudia // Jurnal Psikologi. – 2025. – №14(3). – P. 367-375. – <http://dx.doi.org/10.30872/psikostudia.v14i3>
2. Düzenli H. A systematic review of educational suggestions on generation Z in the context of distance education // Journal of Educational Technology & Online Learning. – 2021. – №4(4). – P. 896-912. – <http://doi.org/10.31681/jetol.1016512>
3. Goran Ćurić, Nikola Papac, Božana Prusina. The challenges of adapting the education system to the requirements of the inclusion of generation Z in the labor market // Special Issue. – 2025. – №1(2025). – P. 9-25. – DOI: 10.47743/rebs-2025-si-0001
4. Elizabeth (Liz) Frissell. The dangers of social media & the stoic solution // International Journal of Theology, Philosophy and Science. – 2020. – №7. – P. 80-88. – <https://doi.org/10.26520/ijtps.2020.4.7.80-88>
5. Cain C. C., Morgan Bryant A., Buskey C. D., & Meyers Ferguson, Y. Generation Z, Learning Preferences, and Technology: An Academic Technology Framework Based on Enterprise Architecture // The Journal of the Southern Association for Information Systems. – 2022. – №9. – P. 1-14. – <https://doi.org/doi:10.17705/3JSIS.00019>
6. Gabriela Philipp, Łukasz Sułkowski. Generative AI and Generation Z: Redefining Language, Identity, and Communication in the Digital Workplace // Sciendo journal. – 2025. – №1(2025). – P. 9-25. – <https://doi.org/10.2478/doc-2025-0014>
7. Carissa Uhlman. Adapting to Gen Z: A Higher Education Guide [Text] / Research brief. – Inceptia, 2018.– P. 2-9.
8. Brenda van Wyk. New Kids on the Block? Exploring Technological Preferences of a New Generation // European Conference on e-Learning. – 2022. – №21(1). – P. 432-437. – <https://doi.org/10.34190/ecel.21.1.446>
9. Hernando L. Bernal Jr. One step ahead of the game: How to prepare for the next generation of students // Academia Letters.– 2021. – Article 540. – P. 1-4. – <https://doi.org/10.20935/AL540>
10. Małgorzata Pikul, Karolina Ostapińska. Motivation of generation Y and Z representatives in the context of organizational resilience // Scientific papers. – 2025. – Silesian University of Technology Publishing House. Organization and management series №237. – P. 491-507. – <http://dx.doi.org/10.29119/1641-3466.2025.237.28>
11. Andrea Tick. Research on the Digital Learning and E-learning Behaviour and Habits of the Early Z Generation // INES. 22nd IEEE International Conference on Intelligent Engineering Systems. – 2018. – №21(1). – P. 33-38. – DOI:10.1109/INES.2018.8523906
12. Hala Shaker Hammad. Teaching the Digital Natives: Examining the Learning Needs and Preferences of Gen Z Learners in Higher Education // Transcultural Journal of Humanities and Social studies. – 2025. – Volume 6 (2). – P. 214-242. –<https://tjhss.journals.ekb.eg/>
13. Szabó Csilla Marianna, Bartal Orsolya, Nagy Bálint. The Methods and IT-Tools Used in Higher Education Assessed in the Characteristics and Attitude of Gen Z // Acta Polytechnica Hungarica. – 2021. – Volume 18 (1). – P. 121-140. – ISSN 1785-8860

Н. Исина, А. Ержуманова, Г. Байдельдинова

ҚАЗАҚСТАНДАҒЫ Z БУЫНЫ СТУДЕНТТЕРІНІҢ АҒЫЛШЫН ТІЛІН ШЕТ ТІЛІ РЕТІНДЕ МЕҢГЕРУДЕГІ ҚАЛАУЛАРЫ: ОНЛАЙН САУАЛНАМАҒА НЕГІЗДЕЛГЕН ЗЕРТТЕУ

Бұл зерттеу Қазақстандық Z буыны студенттерінің (17-18 жас аралығындағы бірінші курс студенттері) ағылшын тілін шет тілі ретінде оқуда қалауларын талдайды. Бұл зерттеудің мақсаты – Z ұрпағының өкілдері, бірінші курс студенттерінің (17-18 жас) ағылшын тілін шет тілі ретінде оқудың таңдаулы әдістері туралы пікірлерін, сондай-ақ олардың қызығушылығы мен білім беру процесіне қатысуына әсер ететін факторларды анықтау. Зерттеу жұмысының эмпирикалық негізі: сипаттамалық статистика және ашық жауаптардың сапалық талдауы арқылы талданған 93 студенттің онлайн сауалнамасы. Студенттердің ағылшын тілін оқуға деген қызығушылығына және сабақтағы белсенділігіне әлеуметтік желілердің, интерактивті/визуалды құралдардың, жасанды интеллекттің әсері, сондай-ақ оқу мінез-құлқының сипаттамаларына – уақытты басқару мен зейін қою және кері байланысқа ерекше назар аударылады. Нәтижелер студенттердің көрнекі және интерактивті құралдарға, әлеуметтік желілерге және сандық құралдарға деген көзқарасын талдайды: студенттер қысқа тапсырмаларды және жылдам кері байланысты қалайды, бұл олардың динамикалық оқу стиліне өте жақсы сәйкес келеді. Жасанды интеллектке деген көзқарастары аралас – олар оның артықшылықтарын мойындайды, бірақ өздерінің дербестігі мен сыни ойлау қабілеттерінің төмендеу мүмкіндігінен қорқады. Қорытындыда ағылшын тілін оқыту үрдісін және бағдарламасын «зумерлардың» сандық әдеттеріне бейімдеу уақыты келді деп атап көрсетілген. Оқу бағдарламаларына арналған ұсыныстар мен технологияны тілдік білім беруге интеграциялау практикалық құндылық болып табылады.

Түйін сөздер: Z буыны, тілдік оқыту, студенттердің қалауы, Қазақстандық университеттер, тілдік оқытудағы әлеуметтік желілер, білім берудегі ЖИ.

Н. Исина, А. Ержуманова, Г. Байдельдинова

ПРЕДПОЧТЕНИЯ СТУДЕНТОВ ПОКОЛЕНИЯ Z В ИЗУЧЕНИИ АНГЛИЙСКОГО ЯЗЫКА КАК ИНОСТРАННОГО В КАЗАХСТАНЕ: ИССЛЕДОВАНИЕ НА ОСНОВЕ ОНЛАЙН-ОПРОСА

В данном исследовании анализируются предпочтения студентов поколения Z (студенты первого курса в возрасте 17-18 лет) в отношении изучения английского языка как иностранного в вузах Казахстана. Цель исследования – изучить мнения студентов первого курса (в возрасте 17-18 лет), представителей поколения Z, об их предпочтительных методах изучения английского языка как иностранного, а также определить факторы, влияющие на их интерес и вовлеченность в учебный процесс. Эмпирическая основа исследования: онлайн-опрос среди 93 студентов, проанализированный с помощью описательной статистики и качественного анализа открытых ответов. Особое внимание уделяется интересу и академической вовлеченности, роли социальных сетей, предпочтениям в отношении интерактивных/визуальных форматов, отношению к ИИ-инструментам, а также особенностям поведения в процессе обучения – от планирования времени и концентрации внимания до ожиданий обратной связи. Результаты демонстрируют интерес к визуальным и интерактивным методам, социальным сетям и цифровым инструментам: студенты предпочитают короткие задания и быструю обратную связь, что идеально соответствует их динамичному стилю обучения. Отношение к ИИ неоднозначно – студенты признают его преимущества, но опасаются потери автономии и утраты критического мышления. В выводах подчеркивается, что необходимо адаптировать обучение английскому языку к цифровым привычкам поколения «зумеров». Практическая ценность исследования заключается в рекомендациях по разработке учебных программ и интеграции технологий в языковое образование.

Ключевые слова: поколение Z, языковое обучение, предпочтения студентов, университеты Казахстана, социальные сети в языковом обучении, ИИ в образовании.

References

1. Nugroho, Y. A. W., Yanti, B. E. D., & Haryanto, F. (2025). A systematic literature review on workplace expectations and behavioral characteristics of Generation Z employees. *Psikostudia: Jurnal Psikologi*, 14(3), 367-375. <http://dx.doi.org/10.30872/psikostudia.v14i3>
2. Düzenli, H. (2021). A systematic review of educational suggestions on generation Z in the context of distance education. *Journal of Educational Technology & Online Learning*, 4(4), 896-912. <http://doi.org/10.31681/jetol.1016512>
3. Ćurić, G., Papac, N., & Prusina, B. (2025). The challenges of adapting the education system to the requirements of the inclusion of generation Z in the labor market. *Review of Economic and Business Studies*, 18(1), 9-25. DOI: 10.47743/rebs-2025-si-0001
4. Frissell, E. (2020). The dangers of social media & the stoic solution. *International Journal of Theology, Philosophy and Science*, (7), 80-88. <https://doi.org/10.26520/ijtps.2020.4.7.80-88>
5. Cain, C. C., Morgan Bryant, A., Buskey, C. D., & Meyers Ferguson, Y. (2022). Generation Z, learning preferences, and technology: An academic technology framework based on enterprise architecture. *The Journal of the Southern Association for Information Systems*, 9(1), 1-14. <https://doi.org/doi:10.17705/3JSIS.00019>
6. Philipp, G., & Sułkowski, Ł. (2025). Generative AI and Generation Z: Redefining language, identity, and communication in the digital workplace. *Discourse and Communication for Sustainable Education*, 16(1), 9-25. <https://doi.org/10.2478/doc-2025-0014>
7. Uhlman, C. (2018). *Adapting to Gen Z: A higher education guide* (Research Brief). Inceptia, 2-9.
8. van Wyk, B. (2022). New kids on the block? Exploring technological preferences of a new generation. *Proceedings of the European Conference on e-Learning*, 21(1), 432-437. <https://doi.org/10.34190/ecel.21.1.446>
9. Bernal, H. L., Jr. (2021). One step ahead of the game: How to prepare for the next generation of students. *Academia Letters*, Article 540, 1-4. <https://doi.org/10.20935/AL540>
10. Pikul, M., & Ostapińska, K. (2025). Motivation of generation Y and Z representatives in the context of organizational resilience. *Scientific Papers of Silesian University of Technology. Organization and Management Series*, (237), 491-507. <http://dx.doi.org/10.29119/1641-3466.2025.237.28>
11. Tick, A. (2018). Research on the digital learning and e-learning behaviour and habits of the early Z generation. In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)* (pp. 33-38). IEEE. DOI:10.1109/INES.2018.8523906
12. Hammad, H. S. (2025). Teaching the digital natives: Examining the learning needs and preferences of Gen Z learners in higher education. *Transcultural Journal of Humanities and Social Sciences*, 6(2), 214-242. <https://tjhss.journals.ekb.eg/>
13. Szabó, C. M., Bartal, O., & Nagy, B. (2021). The methods and IT-tools used in higher education assessed in the characteristics and attitude of Gen Z. *Acta Polytechnica Hungarica*, 18(1), 121-140.

Information about authors:

Issina Nazerke (corresponding author) – Master of Pedagogical Sciences, Lecturer, Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan, e-mail: nazerkeissina01@gmail.com

Yerzhumanova Aziza – Master of Philological Sciences, Lecturer, Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan, e-mail: aziza.yerzhumanova06@gmail.com

Baideldinova Gulmira – PhD, Assistant Professor, Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan, e-mail: baydeldinova76@mail.ru

Авторлар туралы мәлімет:

Исина Назерке Темірқабылқызы (автор-корреспондент) – педагогика ғылымдарының магистрі, оқытушы, Әбілқас Сағынов атындағы Қарағанды техникалық университеті, Қарағанды, Қазақстан, e-mail: nazerkeissina01@gmail.com

Ержуманова Азиза Берикбаевна – филология ғылымдарының магистрі, оқытушы, Әбілқас Сағынов атындағы Қарағанды техникалық университеті, Қарағанды, Қазақстан, e-mail: aziza.yerzhumanova06@gmail.com

Байдельдинова Гульмира Муратовна – PhD, профессор ассистенті, Әбілқас Сағынов атындағы Қарағанды техникалық университеті, Қарағанды, Қазақстан, e-mail: baydeldinova76@mail.ru

Сведения об авторах:

Исина Назерке Темірқабылқызы (автор-корреспондент) - магистр педагогических наук, преподаватель, Карагандинский технический университет имени Абылкаса Сагинова, Караганда, Казахстан, e-mail: nazerkeissina01@gmail.com

Ержуманова Азиза Берикбаевна – магистр филологических наук, преподаватель, Карагандинский технический университет имени Абылкаса Сагинова, Караганда, Казахстан, e-mail: aziza.yerzhumanova06@gmail.com

Байдельдинова Гульмира Муратовна – PhD, ассистент профессора, Карагандинский технический университет имени Абылкаса Сагинова, Караганда, Казахстан, e-mail: baydeldinova76@mail.ru

S. Kastayev¹, A. Tursynbayeva², N. Sartayeva^{3*}, S. Mambetaliyev⁴, A. Bostanova⁵

¹Central Asian Innovation University, Shymkent, Kazakhstan

²M. Auezov South Kazakhstan University, Shymkent, Kazakhstan

³Kazakh National Pedagogical University named after Abai, Almaty, Kazakhstan

^{4,5}Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan

*e-mail: nazgul_88s@mail.ru

¹ORCID 0000-0003-3144-8310, ²ORCID 0009-0001-3055-8617, ³ORCID 0000-0001-5123-5553

⁴ORCID 0009-0004-5407-0191, ⁵ORCID 0000-0001-8035-9854

DIAGNOSTICS OF THE FORMATION OF ORGANIZATIONAL SKILLS OF FUTURE TEACHERS

The article discusses the problem of the development of organizational skills of future teachers. The relevance of the research problem was determined based on the regulatory documents, a review of scientific and pedagogical works related to the research topic was conducted, as well as the issue of forming organizational skills was analysed. Motivational, meaningful, activity-based and reflexive- evaluative components, criteria and indicators of the formation of organizational skills of future teachers are identified carefully as well as skills, formation, high, medium and low levels are described. A scheme of meaningful characteristics of the organizational skills of future teachers is drawn up by giving explanations for each organizational and managerial area. At the stage of the experiment, quantitative data are obtained, diagnostic methods for organizational knowledge, emotional value sphere, and skills are used. As a result of the diagnosis, it was found that the state of formation of organizational and managerial directions among modern students is at average and low level, especially in the complete absence of a high level of activity component. Due to the fact that the organization and the future is professionally important for teachers, special attention was paid to the development of a program for their formation in the educational process of a pedagogical university.

Keywords: future teacher, organization, knowledge, skills, abilities, management, emotional- value sphere, diagnostics.

Introduction

In the context of renewal of all spheres of society, the teacher's function is becoming much more complicated. A modern school needs to train professionals in higher education institutions who are able to model and systematically improve their professional activities, quickly and effectively introduce scientific achievements and innovative practices into the pedagogical process.

The training of future teachers involves the formation of their professional competencies i.e. organizational, leadership, creative, creative, meta-subject, etc. It is known that the formation of skills serves to strengthen the modern model of the student. The main direction of modernization of education in the country is characterised by an orientation towards openness, universal values, integration of the learning and upbringing process which requires the development of professional knowledge and educational activities, organizational skills, and personal qualities in the training of future teachers.

The third mission of the university was the concept of development of higher education and science of the Republic of Kazakhstan for 2023-2029: "a comprehensive system of student education will be created that meets the goals, content and achieved results of the state policy in the field of education and upbringing of youth" [1]. One of the possible ways to improve education in a modern university in this regard is to improve the organizational skills of future teachers in educational activities. This, in turn, will significantly increase educational activities. It is argued that the organizational skills of future teachers have a great influence on the implementation of subjective program ideas in the development of the university's third mission.

The solution of the tasks set for the training of future teachers is reflected in the regulatory documents, including the state mandatory standards of education of the Republic of Kazakhstan, as "achieving a balance between the academic and practical orientation of the educational content" [2].

Here organizational skills of future teachers play a high role in increasing the importance of practical and academic orientation.

Research in the field of training future teachers in a university setting shows the need for future teachers organizational skills to form an educational environment, organise educational activities, and commit to mass training of talented teachers at the university. Increasing the organizational skills of future teachers makes it possible to solve the problems of organising the educational activities of students in the future in general education institutions, to organise a productive learning process based on the acquired skills.

The ideas of future teachers about the management and organization of the educational process were examined in a joint scientific study by Kaufman and Moss. The results showed that respondents' answers were strongly influenced by the way the questions were presented and varied depending on their structure and students' interpretation. The study also revealed contradictions between students' views on management and organization and their theoretical orientations, as well as insufficient attention to their own professional development. The authors concluded that students need to deepen their understanding of organization and management and receive support in bridging the gap between theory and practice [3]. These findings point to the need for purposeful development of organizational skills among future teachers and emphasize the relevance of the research problem.

This issue has also been addressed in domestic research. Kastaev and Grinshkun examined the organizational skills of future teachers and identified their main components, emphasizing the importance of their development in teacher education [4]. Yertileuova et al. further examined the strategic and organizational literacy of future educators in relation to their managerial potential [5].

Krepf and König focused on lesson planning as an important aspect of organizing learning activities and examined lesson structuring as a component of teachers' professional planning competence [6]. Their study emphasizes the importance of planning lesson structure and organizing instruction as interconnected aspects of teaching practice. Gidalevich and Bigman also highlighted organized behavior among the professional characteristics of future teachers, including time management, task completion, and adherence to established requirements [7]. Furthermore, Agranovich et al. investigated the self-organization of prospective primary school teachers through time-management technology and emphasized the importance of planning and regulating academic activities [8]. Together, these studies highlight organizational skills as an important part of future teachers' professional preparation. The main future of educational activity can be distinguished as "self development", "self educational activity" and its main result is the development of a person's skills, orientation, and personal qualities. In this regard, the teacher must have certain organizational skills necessary for the organization of purposeful educational activities.

The process of preparing a future teacher for professional organizational activity is based on the acquisition of knowledge, specific skills, the assimilation of specific behavioral patterns, etc. The implementation of educational activities depends on pedagogical competencies. Skill is a complex psychological and pedagogical concept. In pedagogical research, the definition of this concept is given as follows: "skill is the effective performance of actions in accordance with set goals and set conditions. It is possible to state a high level of abilities of a person who quickly and efficiently mastered the skill" [9].

The university lays the foundation for a high level of not only for general but also specialised professional skills of a person, formed primarily on the basis of general skills, the ease and speed of mastering professional skills by a student.

In order to fulfill educational tasks, B. D. Krasovsky understood organizational business as readiness to implement a system of practical actions based on knowledge and experience aimed at regulating the activities of the Team [10].

It is known that the main activity of a teacher is planning, organising, the educational process and evaluating it. In other words, there is a need for instruction, that is for the organisation of pedagogical activity. Pedagogical activity serves to implement the functions of teaching, evaluating, and fostering the development of students' knowledge [11].

V.G Suranov formulated the definition of organizational competence as follows: “we mean the teacher’s ability and willingness to carry out organisational activities based on acquired skills, knowledge, and accumulated experience and developed activities” [12].

According to S.K Kastaev, a teacher's organisational skills refer to the ability to be conscious, effectively and timely carry out practical actions aimed at planning, organising, and regulating the activities of a group of students in order to solve pedagogical tasks in accordance with established values (as a component of training) [13].

Thus, it was established that the structural components of the organisational skills of teachers in the organization of the learning process in schools from the point of structural substantive, professionographic, functional, conceptual, competency pedestals consist of knowledge, activity (skill) and, accordingly, qualitative personal qualities. Each of the listed components has its own content and is clarified by skill types according to the phrases.

In the course of studying the topic of the conducted research and the written scientific literature, we were convinced of the insufficiency of diagnostic studies confirmed by specific quantitative data, and we noticed that the problem of the formation of organizational skills of future teachers was not fully considered.

The purpose of the study is to identify the current state of organisational skills of future teachers, identify components and directions, and diagnose the level of formation.

The main provisions:

Skill is the ability to carry out some actions competently.

Organizational skill is the organizational ability that includes the activities of design, planning and implementation.

The importance of the formation of organizational skills of a teacher is to ensure the effectiveness of the teacher's work, to make the learning process productive.

In the course of the study, the essence of the concept of organizational skill is considered, the importance of the formation of organizational skills of future teachers, directions are determined, and the diagnosis of the state of formation is carried out.

Materials and research methods

Since the research problem is related to the relevant topic in the higher education system, it requires a comprehensive analysis from a pedagogical and methodological point of view. In the course of the research, the analysis of the consideration of key concepts in scientific research using the theoretical method was carried out and the level of developing organizational skills of future teachers was diagnosed using empirical research. Based on the analysis of the literature on the issue under consideration and in accordance with the specifics of organizational skills, we determine the components, criteria and indicators for the formation of organisational skills (Table 1).

Table 1 - Components, criteria, indicators for the formation of organisational skills of future teachers

Components	Criteria	Indicators
Motivational	Activity, awareness, curiosity, aspiration	- the desire to learn the essence of the learning process and the activity to do it; - conscious awareness of the theoretical foundations of training activities; -the desire to fully implement the skills of organizing training activities.

Content	In accordance with the implementation of organizational skills	- knowledge of the essence of educational activities; - knowledge of the theoretical foundations of the learning process; -knowledge of a detailed description of the organization of training activities.
Activity	Activities in accordance with the implementation of organizational skills	- development of necessary documents (planning) for the training process; - Organization of the educational process; -Organization of assessment in training activities.
Reflexive-evaluative	Analysis and evaluation of actions in accordance with the implementation of their organizational skills	- analysis of organized training activities; -control and evaluation of organized training activities; -make adjustments to organizational activities for training.

In accordance with theoretical research, we have identified the formed level of organisational skills of teachers as follows:

High level. At this level. Future teachers have a high organisation of educational activities, very high knowledge and skills to organise training in structuring basic concepts and terms in educational materials and set out in detail the methodology of organising training. He is fully proficient in the educational material and appropriately selects and evaluates content in accordance with a specific topic and activity.

The medium level. At this level, future teachers better organise educational activities, have knowledge of structuring basic concepts and terms in educational materials, the ability to organise learning and describe the methods of organising learning. The demonstration of educational material and the selection, evaluation in accordance with a specific topic, activity cannot explain the full value.

Low level. At this level students have unsatisfactory skills in organising educational activities, they are able to structure. Basic concepts and terms in educational materials cannot thoroughly explain at the knowledge level characterise the course of instructional abilities and organizations, and the methodology with errors. The selection and evaluation of educational material and regarding the specific topic, activity is performed with small errors.

Thus, the problem of the formation of organizational skills of future teachers can be called a complex structure, which includes such activities as design, planning, analysis, evaluation, orientation, reflection, consisting of motivational, substantive, action and reflexive-evaluative components. Its content description is given in Figure 1.



Figure 1 - Content description of organizational skills of future teachers

In compiling this picture, we intended to cover the main types of activities and professional qualities that characterise organisational skills. The following components are presented as basic concepts in the figure: organisational education, organisational skill, and emotional- value sphere. Based on the individual determination of the level of formation of each of them, the level of development of organisational skills of future teachers is determined.

One of the components of the organisational skills of future teachers is the organisational skills of future teachers is organising knowledge. It corresponds to the content component. The students were assigned diagnostic tasks consisting of 10 open- ended questions aimed at identifying the formation of organisational knowledge. The survey was conducted on Google drive The following issue were covered:

1. Educational activities of students -.... .
2. Are there any (substantive) features of the concepts of" educational activity "and" teaching activity"? If so, write it down.
3. How do you explain what we mean by the teacher's teaching activity?
4. How do you understand the concept of" pedagogical activity"?
5. highlight the types of pedagogical activity of the teacher.
6. How do you interpret the concept of" Organization of teaching activities of a teacher"?
7. write down the stages of the teacher's implementation of the organizational skills of teaching activities.
8. describe the content of the teacher's design (design) activity in the implementation of training.
9. describe the content of organizational activities in the implementation of educational activities of students.
10. describe the content of the reflexive-analytical activity of the teacher in the learning process.

An assessment of the implementation of the task with 1 (low), 2 (medium) and 3 (high) is established, depending on the correctness and completeness. In the overall results, it was planned to mark 25-30 points - high, 16-24 points- medium and 10-15 points - low level.

Organisational skill: specific organizational abilities and skills of the teacher were determined by practical control. Types of organizational skills: analytical-project, communicative-developmental, reflective. To determine the level of formation of organizational skills, the pedagogical task "drawing up a plan of educational activities of students on a given topic" was set."He himself was judged by the type of skill. Interpretation, analysis and design: determining the purpose of the task, identifying possible difficulties; establishing stages, steps to complete the task, drawing up a plan; communicative and Developmental: Organization of students ' activities, assignment of tasks, instruction, distribution of work according to the forms of study, preparation of tasks that increase the activity of students; reflexive: planning feedback forms and reflections. The organization of work for each indicator is estimated at 0-2 points, the total number of points is 6. of these: 5-6-High, 3-4 - medium, 0-2-low.

Components of the emotional and value sphere include student perception; responsiveness; choice of values; value formation; manifestation of values in action. These components work primarily on two components of organizational skills: motivational and reflective-evaluative. To determine the level of organizational skills training of future teachers in terms of the components of the emotional and value sphere, a survey of 10 statements was proposed, which was evaluated on the Likert scale (1-totally disagree; 2-disagree; 3-I can't say for sure; 4-agree; 5 - Completely agree). The result was refined according to high (3850 points), average (2437 points) and low (1023 points) levels.

Results and discussion

The pedagogical experiment involved 3 3rd year students of the primary education program of the Kazakh National Pedagogical University named after Abai (total number - 81 students). In order

to determine the level of formation of organisational skills of future teachers, a determination experiment was carried out on the research work.

In the course of the detection experiment, the initial level of organisational skills of the respondents was determined. To this end, we divided the respondents into two groups; control (41 students) and experimental (40 students). We applied an individual diagnostic method to each organisational skill component. The level of formation of organisational knowledge was identified by the diagnostic task, the level of formation of skills was described by the practical task, the level of formation of organizational knowledge was determined by the diagnostic task, the level of formation of skills was determined by the practical task, the level of organizational business of future teachers by the components of the emotional and value sphere was determined by the questionnaire. The analysis of the answers to the diagnostic task given in order to predispose the level of organizational knowledge gave rise to the following conclusions, i.e., in the overwhelming majority of students:

- describes the educational activities of students in a general manner, does not give an exhaustive and clear answer;
- does not fully distinguish the substantive features of the concepts of "educational activity" and "training activity". It cannot describe that educational activities are aimed at organizing educational activities and that educational activities are carried out independently by the student, the basis of which is motivation;
- the teacher cannot fully explain the fact educational activities are carried out through a unity of actions aimed at the purpose and results of learning, and through the stages of this training related to design before training ;
- the concept of «pedagogical activity» poorly characterises what consists of educational, educational and instructional activities;
- there is a confusion of the teacher with the types of pedagogical activity, its functions;
- cannot fully explain the concepts of "teacher's organizational skills" "teacher's organizational skills in educational activities" ;
- does not fully reflect the stages of realization of the teacher's ability to organize educational activities;
- knocks down the assembly (design) activity of the teacher when teaching with other skills;
- to disclose the content of students' organizational activities in the implementation of educational activities;;
- the description of the content of the reflexive and analytical activity of the teacher in the learning process is not exhaustive.

The level of students' organizational knowledge can be seen in Table 2 and Figure 2.

Table 2 - The level of manifestation of organizational knowledge (by diagnostic stage)

Formation level	Results	
	EG, % (40)	CG, % (41)
High	5	4,9
Medium	25	21,9
Low	70	73,2

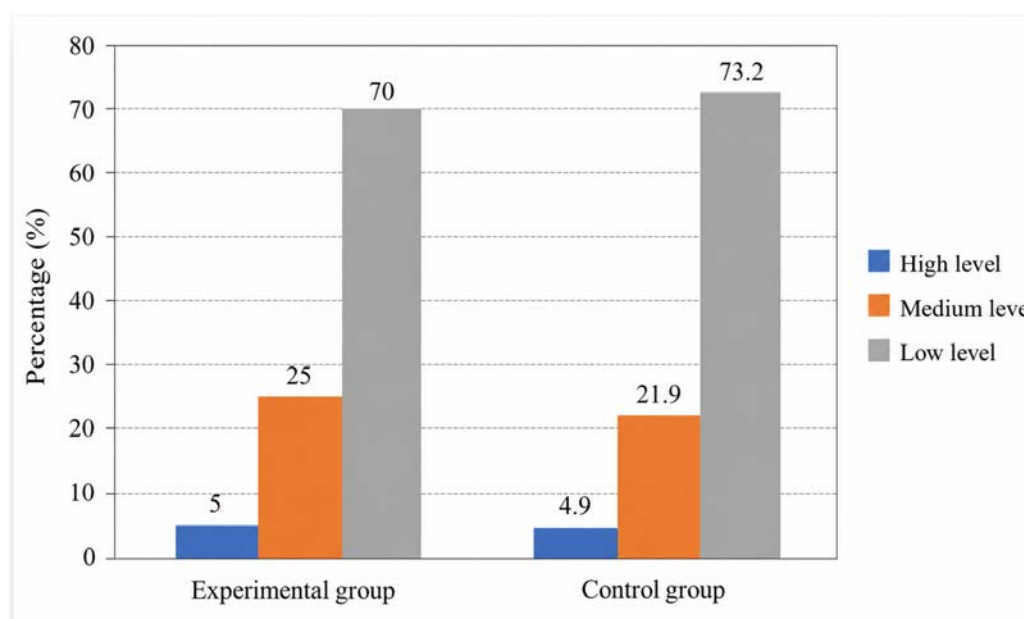


Figure 2 - The level of manifestation of organizational knowledge (by diagnostic stage)

Following that components of organisational skills were identified. The results of the implementation of the pedagogical task by future teachers we have shown in table 3.

Table 3 - indicators of the level of formation of organizational skills (by the period of determination)

Organisational skills	Levels					
	Low		Medium		High	
	EG, % (40)	CG, % (41)	EG, % (40)	CG, % (41)	EG, % (40)	CG, % (41)
Analytical and design	74,1	75,7	25,9	24,3	-	-
Communicative and developmental	54,6	56,2	45,4	43,8	-	-
Reflexive	53,4	53,7	46,6	46,3	-	-
The average value	60,7	61,9	39,3	38,1	-	-

The table shows that organisational skills are divided into 2 levels- low and medium. However, the majority of respondents in the control (60.7%) and experimental (61.9%) groups showed a low level in them. A small part of the respondents (25.9%) of EG and (24.3%) of CG managed to show independence in analysing tasks and finding ways to solve them. Partially, using their new untested ideas, they were able to act strictly according to a given algorithm. The formation of analytical skills of such students was of an average level.

74.1% and 75.7% of the participants in the experiment showed a low degree of analytical and design skills. This was manifested in the incorrect construction of the activities of future teachers in planning , analysing the educational activities of students, identifying and solving difficulties. The vast majority of test students did not take into account the observance of consistency and in the construction of steps to complete the tasks. The ability to design illustrates that the process of setting goals and objectives, when choosing the content and means of organisational activity, students' actions are carried out strictly according to the model, without taking into account the

individual characteristics of students, using well known forecasting methods. The structure of these skills includes actions to develop a lesson plan and prepare a summary. It was determined that they were only copying, and showed a low level. Since this is the most important skill in the professional activity of a teacher, it was concluded that his indicator requires improvement measures.

The indicator for the communicative and developmental skills of 54.6% and 56.2% was revealed under the influence of shortcomings made by future teachers in the organization of students' activities and the allocation of tasks, failure to follow the sequence of specific steps in issuing the necessary instructions, failure to work in a group form and the compilation of developmental tasks. Thus, it has been established that the creative organization of the student team for educational activities is carried out by the respondents strictly according to the model, under the guidance of teachers using certain methods of activity. Only 45.4% and 43.8% of respondents from each group tried to correctly complete the specified tasks, but the organizational skills according to the criteria were not fully realized correctly. That's why they got an average score.

The analysis of the results showed that 53.4% of the students in the experimental group and 53.7% of the respondents in the control group showed insufficient ability to establish feedback and organize reflexive activities. According to this indicator, only 46.6% and 46.3% of future teachers made their actions close to the required level, only mistakes were made in organizing feedback and reflection, did not strive for innovation, thereby did not establish a trajectory of professional development. This suggests that reflexive skills are not fully developed.

In terms of mastering the components of the emotional and value sphere, future teachers showed an average and low level. Failure to receive a high-level response indicates the students' lack of interest in organizational activities in professional and pedagogical activities, internal motivation, and the predominance of only external influences. The results of the diagnostic method performed for this type of activity are presented in Table 4 and Figure 3.

Table 4 - The level of manifestation of the relationship of the emotional and value sphere

Formation level	Results	
	EG, % (40)	CG, % (41)
High	5	7,3
Medium	45	46,3
Low	50	46,4

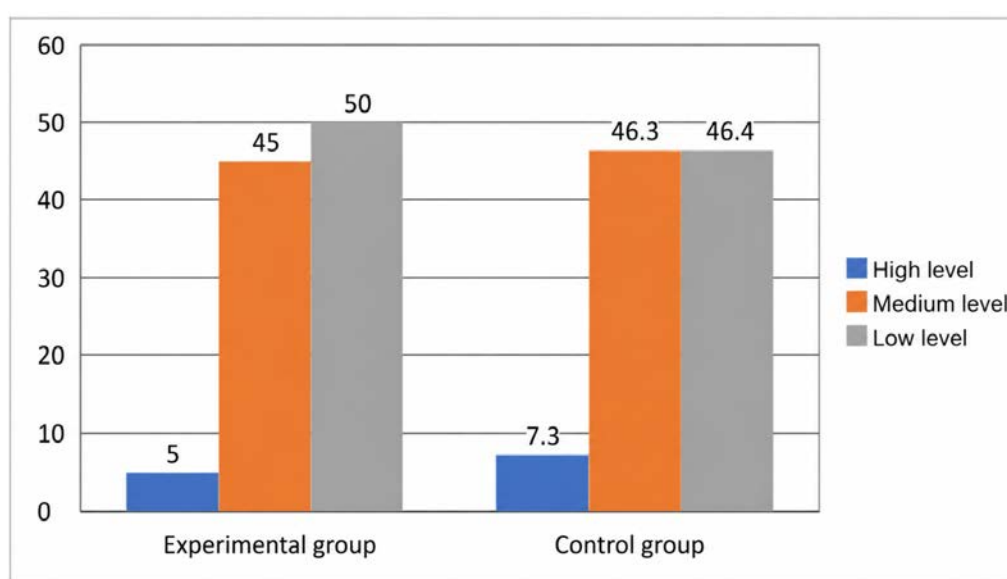


Figure 3 - The level of manifestation of the relationship of the emotional and value sphere

A comparison of the experimental and control groups in terms of the emotional- value sphere component of the acquired organisational skills and the level of formation of pedagogical organisational experience in general shows their uniformity.

Thus, the integration of the results obtained through the diagnostic methods applied at the detection stage, calculating that average values, and their graphical representation, made it possible to determine the actual level of organisational skills among future teachers in the contemporary educational context (Table 5, Figure 4)

Table 5 - The cumulative level of formation of organizational skills of future teachers (by the stage of determination)

Formation level	Results	
	EG, % (40)	CG, % (41)
High	3,3	4,1
Medium	36,4	35,4
Low	60,3	60,5

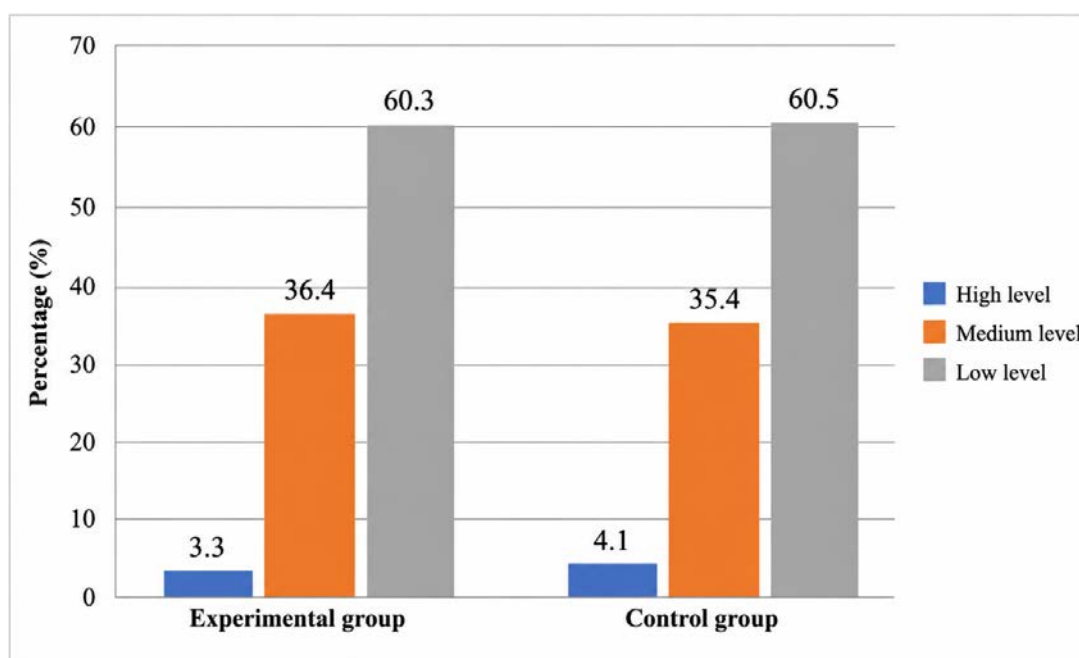


Figure 4 - Diagram of the formation of organisational skills of future teachers

Thus, at the Diagnostic stage, the initial level of formation of organisational skills in the experimental and control group was determined. To calculate the statistical significance of the differences between the results the Pearson's χ^2 criterion was applied.

Pearson's χ^2 value is intended to test the hypothesis of independence of variables or compliance of observed data with theoretical distributions.

χ^2 - value of chi-square statistics

O_i - empirical frequency

E_i - theoretical frequency

The calculation results showed that $\chi^2_{emp} = 0.33$. Given the degree of freedom $df = 2$, the critical value was $\chi^2_{crit} = 5.99$. Since $\chi^2_{emp} < \chi^2_{crit}$, there is no statistically significant difference between the results of the experimental and control groups. This indicates that the groups were comparable in terms of their initial level of development and that the study could be further continued under comparable research conditions.

In this way, as a result of the study, it was found that the level of formation of organizational skills of future teachers requires additional work in specially organised conditions. This coincides with the conclusions of an international study. Akbulut and Karabay considered the problem of improving the skills of future teachers creating the content of educational material. It was identified by the experiment that students were not able to adequately conduct the method of selection organisation - integration of information necessary for the organisation of educational content. In accordance with this situation, it was concluded that in order to accustom future teachers to organisational work, well structured support from teachers of higher educational institutions is needed [14]. Moreover, in the study of Street et al., the essence of the organisational and emotional abilities of future teachers is considered. Quantitative data from the experiment showed that the respondents had a low level of organisational competence. It was concluded that one of the reasons contributing to this might be the priority of the cognitive load of the future teacher and the need to develop organisational competence as a center of professional Development [15]. The results of the research leads to the same conclusion.

The formation of organisational skills at a high level is of great importance for the professional development and self- realization of the future teacher, and the effectiveness of the learning process as a whole. To improve the results, it is necessary to identify objective factors that hinder the preparation of a future teacher. In particular, the imperfection of educational programs of preparation for organisational educational activities has been revealed. Most programs do not provide for the development of organisational skills and actions among students, in addition, the system for the formation of these skills during extracurricular activities is not sufficiently developed. Their creation is carried out mainly in practical activities, students do not always perform it, so it does not even give the desired result. Such randomly formed skills are characterised by fragments and limited transference, narrowing the scope of application.

The formation and development of organisational skills become purposeful if students discover the sequence of actions that make up certain skills, and understand the value of mastering them. An important condition for the purposeful development of skills is the creation of a special program for students to master organisational skills.

The organisational and managerial competencies constitute one of the central components in the process of preparing future teachers for educational and instructional activities. Since, organisational and managerial skills are professionally significant for future teachers, special attention should be devoted to developing a program for their formation within the educational process of a pedagogical university. The purposeful development of future teachers' organisational skills is carried out in accordance with the principle of "teaching through development" whereby acquired skills are organised and applied in new ways, subsequently serving as a foundation for the further enhancement of organisational competencies.

Conclusion

Considering the problem of formation of organisational skills of future teachers, it was found that it is a complex integrative structure. The full fledged formation, systematic application and continuous improvement of these skills can pave the way for the effectiveness of the future teacher's professional activities, the professional personal development, and search for solutions to pressing the problems in the education system. In the course of the research, such diagnostic areas of organisational skills as organisational knowledge, skills and emotional value spheres were identified. It has been established that due to the low effectiveness of diagnosing the state of formation of these abilities lines among modern students, especially with the complete lack of a high level of activity component, this issue requires a comprehensive methodological study and the implementation of formative measures.

Based on the data of the scientific and pedagogical analysis of the problem and the results of the research, the following scientifically sound recommendations have been developed:

1. The inclusion in the university's work programs of tasks on the formation of organisational skills of future teachers;

2. Organisation of work in the educational process according to skill types of analytical, project, communicative, developmental, reflective orientation;

3. Inclusion in the long term plan or formative research aimed at determining effective conditions and a methodological system for the formation of organisational skills of future teachers.

The components, levels, structural diagnostic directions and empirical data obtained from it, identified as a result of the diagnosis of the formation of organisational skills of future teachers, serve as the basis for the development of a scientific and methodological system for the formation of organisational skills. In the future, it is necessary to continue studying in the following areas; formation of organisational skills of the educational activity of the future teacher on the basis of pedagogical tasks; formation of organisational skills of the future teacher on the basis of game technologies; improving the methodological competence of future teachers based on game technologies, etc.

References

1. Қазақстан Республикасында жоғары білімді және ғылымды дамытудың 2023 - 2029 жылдарға арналған тұжырымдамасы. Қазақстан Республикасы Үкіметінің 2023 жылғы 28 наурыздағы № 248 қаулысы. <https://adilet.zan.kz/kaz/docs/P2300000248>

2. Жоғары және жоғары оқу орнынан кейінгі білім берудің мемлекеттік жалпыға міндетті стандарты. Қазақстан Республикасы Ғылым және жоғары білім министрінің 2022 жылғы 20 шілдедегі, №2 бұйрығы. <https://adilet.zan.kz/kaz/docs/V2200028916>

3. Kaufman D., Moss D.M. A new look at preservice teachers' conceptions of classroom management and organization: uncovering complexity and dissonance. //The Teacher Educator, 2010. - 45(2). - 118-136 <https://doi.org/10.1080/08878731003623669>

4. Krepf, M., König, J. Structuring lessons as an aspect of pre-service teachers' planning competence: A scaling-up analysis. //Z Erziehungswiss, 2022. - 25. - 917-946, <https://doi.org/10.1007/s11618-022-01125-9>

5. Халмурадов Т.Н. Формирование интеллектуальных умений у учащихся 5-7 классов в процессе трудового обучения: дис. канд. пед. наук. - М., 1992. - 17 с.

6. Красовский Б.Д. Формирование руководителя нового типа. Проблемы теории и практики управления. - М., 1992. - №6. - С. 120.

7. Ярославова Г.Ю. Формирование организаторских умений будущего учителя в условиях студенческого педагогического отряда: дис. канд. пед. наук. - Челябинск, 2007. - 196 с.

8. Суранов В.Г. Формирование организаторской компетенции у студентов педагогического вуза в условиях студенческого клуба. //Вестник ЧГПУ им. И.Я. Яковлева. - 2010. - Т. 1, №3(67). - С. 168-174. <https://cyberleninka.ru/article/n/formirovanie-organizatorskoy-kompetentsii-u-studentov-pedagogicheskogo-vuza-v-uslovih-studencheskogo-kлубa>

9. Akbulut T., Karabay A. Improving Student Classroom Teachers' Content Organization Skills Through Practicum. //Participatory Educational Research, 2023. - 10(1). - 42-66. <https://doi.org/10.17275/per.23.3.10.1>

10. Street K.E.S., Ertesvåg S.K., Gamlem S.M., Rebni L.M., Solvik R.M., Vaaland G.S. Classroom management for first-year pre-service teachers: What is difficult and why? //Teaching and Teacher Education, 2025. - Volume 168. - 105235. <https://doi.org/10.1016/j.tate.2025.105235>

11. Кастаев С.К. Болашақ мұғалімдердің ұйымдастырушылық іскерліктерін педагогикалық тапсырмалар негізінде қалыптастыру: филос.д-ры. дис. - Түркістан, 2025. - 180 б.

12. Суранов В.Г. Формирование организаторской компетенции у студентов педагогического вуза в условиях студенческого клуба. //Вестник ЧГПУ им. И.Я. Яковлева. - 2010. - Т. 1, №3(67). - С. 168-174. <https://cyberleninka.ru/article/n/formirovanie-organizatorskoy-kompetentsii-u-studentov-pedagogicheskogo-vuza-v-uslovih-studencheskogo-kлубa> 11.06.2026

13. Weyers J., Kramer C., Kaspar K., König J. Measuring pre-service teachers' decision-making in classroom management: A video-based assessment approach // Teaching and Teacher Education. - 2024. - Vol. 138. - Article 104426. <https://doi.org/10.1016/j.tate.2023.104426>

14. Li L., Hu Y., Yang X., Wu M., Tao P., Chen M., Yang C. Enhancing pre-service teachers' classroom management competency in a large class context: the role of fully immersive virtual reality // Humanities and Social Sciences Communications. - 2024. - Vol. 11. - Article 1050. P. 1-14. <https://doi.org/10.1057/s41599-024-03538-9>

15. Kramer C., Strauß S., König J. Student teachers' learning about effective classroom management: efficacy and transfer of a video-based intervention aimed at promoting situation-specific skills and self-efficacy // Zeitschrift für Erziehungswissenschaft. - 2025. - Vol. 28. - P. 1501-1523. <https://doi.org/10.1007/s11618-025-01359-3>

С. Кастаев, А. Турсынбаева, Н. Сартаева, С. Мамбеталиев, А. Бостанова

БОЛАШАҚ ПЕДАГОГТЕРДІҢ ҰЙЫМДАСТЫРУШЫЛЫҚ ІСКЕРЛІКТЕРІН ҚАЛЫПТАСТЫРУДЫҢ ДИАГНОСТИКАСЫ

Мақалада болашақ мұғалімдердің ұйымдастырушылық іскерліктерін дамыту мәселесі қарастырылады. Зерттеу мәселесінің өзектілігі нормативтік құжаттар негізінде айқындалды, зерттеу тақырыбына қатысты ғылыми-педагогикалық еңбектерге шолу жасалды, сондай-ақ ұйымдастырушылық іскерліктерді қалыптастыру мәселесі талданды. Болашақ мұғалімдердің ұйымдастырушылық іскерліктерін қалыптастырудың мотивациялық, мазмұндық, іс-әрекеттік және рефлексивті-бағалаушылық компоненттері, критерийлері мен көрсеткіштері мұқият айқындалып, іскерліктер мен олардың қалыптасу деңгейлері (жоғары, орташа және төмен) сипатталды. Әрбір ұйымдастырушылық-басқарушылық бағытқа түсініктемелер беру арқылы болашақ мұғалімдердің ұйымдастырушылық іскерліктерінің мазмұндық сипаттамаларының сызбасы жасалды. Тәжірибелік кезеңде сандық деректер алынып, ұйымдастырушылық білімді, эмоционалдық-құндылықтық саланы және іскерліктерді бағалауға арналған диагностикалық әдістер қолданылды. Диагностика нәтижесінде заманауи студенттер арасында ұйымдастырушылық-басқарушылық бағыттардың қалыптасу деңгейі орташа және төмен екендігі, әсіресе белсенділік компонентінің жоғары деңгейі мүлдем байқалмайтыны анықталды. Ұйымдастырушылық қызмет пен болашақ кәсіби қызмет мұғалімдер үшін маңызды болғандықтан, педагогикалық жоғары оқу орнындағы оқу-тәрбие процесінде осы бағыттарды қалыптастыруға арналған бағдарламаны әзірлеуге ерекше назар аударылды.

Түйін сөздер: болашақ педагог, ұйымдастыру, білім, іскерлік, қабілеттер, басқару, эмоционалды-құндылықтық сала, диагностика.

С. Кастаев, А. Турсынбаева, Н. Сартаева, С. Мамбеталиев, А. Бостанова

ДИАГНОСТИКА ФОРМИРОВАНИЯ ОРГАНИЗАТОРСКИХ УМЕНИЙ БУДУЩИХ УЧИТЕЛЕЙ

В статье рассматривается проблема развития организаторских умений у будущих педагогов. Актуальность проблемы исследования обоснована на основе нормативных документов; проведен обзор научно-педагогических работ по теме исследования, а также проанализирован вопрос формирования организаторских умений. Детально определены мотивационный, содержательный, деятельностный и рефлексивно-оценочный компоненты, а также критерии и показатели сформированности организационных умений будущих педагогов; описаны сами умения и уровни их сформированности (высокий, средний и низкий). Разработана схема содержательных характеристик организационных умений будущих педагогов с пояснениями по каждой организационно-управленческой области. В ходе эксперимента получены количественные данные и применены методы диагностики организационных знаний, эмоционально-ценностной сферы и умений. По результатам диагностики было установлено, что уровень сформированности организационно-управленческой направленности у современных студентов является средним или низким, причем высокий уровень деятельностного компонента практически отсутствует. Учитывая профессиональную значимость организаторских качеств для будущих педагогов, особое внимание

было уделено разработке программы их формирования в рамках образовательного процесса педагогического вуза.

Ключевые слова: будущий педагог, организация, знание, умение, способность, управление, эмоционально-ценностная сфера, диагностика.

References

1. Government of the Republic of Kazakhstan. (2023). *Qazaqstan Respublikasynda jogary bilimdi jäne gylymdy damytudyn 2023 - 2029 jyldarga arналған tujrymdamasy* [Concept for the development of higher education and science in the Republic of Kazakhstan for 2023–2029]. Decree of the Government of the Republic of Kazakhstan dated March 28, 2023, No. 248. <https://adilet.zan.kz/kaz/docs/P2300000248>
2. Ministry of Science and Higher Education of the Republic of Kazakhstan. (2022). *Jogary jäne jogary oqu ornynan keyingi bilim berudin memlekettik jalgypa mindetti standarty* [State obligatory standard of higher and postgraduate education]. Order of the Minister of Science and Higher Education of the Republic of Kazakhstan dated July 20, 2022, No. 2. <https://adilet.zan.kz/kaz/docs/V2200028916>
3. Kaufman, D., & Moss, D. M. (2010). A new look at preservice teachers' conceptions of classroom management and organization: Uncovering complexity and dissonance. *The Teacher Educator*, 45(2), 118–136. <https://doi.org/10.1080/08878731003623669>
4. Krepf, M., & König, J. (2022). Structuring lessons as an aspect of pre-service teachers' planning competence: A scaling-up analysis. *Z Erziehungswiss*, 25, 917–946. <https://doi.org/10.1007/s11618-022-01125-9>
5. Khalmuradov, T. N. (1992). *Formirovanie intellektualnykh umeniy u uchashchikhsya 5-7 klassov v protsesse trudovogo obucheniya* [Formation of intellectual skills in 5th-7th grade students in the process of labor training] (Candidate dissertation). Moscow, 17 p.
6. Krasovskiy, B. D. (1992). *Formirovanie rukovoditelya novogo tipa* [Formation of a new type of leader]. *Problemy teorii i praktiki upravleniya**, (6), 120.
7. Yaroslavova, G. Yu. (2007). *Formirovanie organizatorskikh umeniy budushchego uchitelya v usloviyakh studencheskogo pedagogicheskogo otryada* [Formation of organizational skills of a future teacher in the conditions of a student pedagogical detachment] (Candidate dissertation). Chelyabinsk, 196 p.
8. Suranov, V. G. (2010). *Formirovanie organizatorskoy kompetentsii u studentov pedagogicheskogo vuza v usloviyakh studencheskogo kluba* [Formation of organizational competence among students of a pedagogical university in the conditions of a student club]. *Vestnik ChGPU im. I.Ya. Yakovleva*, 1(3), 168–174. <https://cyberleninka.ru/article/n/formirovanie-organizatorskoy-kompetentsii-u-studentov-pedagogicheskogo-vuza-v-uslovih-studencheskogo-kluba>
9. Akbulut, T., & Karabay, A. (2023). Improving student classroom teachers' content organization skills through practicum. *Participatory Educational Research*, 10(1), 42–66. <https://doi.org/10.17275/per.23.3.10.1>
10. Street, K. E. S., Ertesvåg, S. K., Gamlem, S. M., Rebni, L. M., Solvik, R. M., & Vaaland, G. S. (2025). Classroom management for first-year pre-service teachers: What is difficult and why? *Teaching and Teacher Education*, 168, 105235. <https://doi.org/10.1016/j.tate.2025.105235>
11. Kastaev, S. K. (2025). *Bolashaq muğalimderdin uymdastyrushylyq iskerliklerin pedagogikalyq tapsyrmalar negizinde qalyptastyru* [Formation of organizational skills of future teachers on the basis of pedagogical tasks] (Doctoral dissertation). Turkistan, 180 p.
12. Suranov, V.G. (2010). *Formirovanie organizatorskoi kompetensii u studentov pedagogicheskogo vuza v uslovih studencheskogo kluba* [Formation of organizational competence among students of a pedagogical university in the conditions of a student club]. *Vestnik CHGPU im. I.İa. İakovleva*, T.1, №3(67), 168-174.
13. Weyers, J., Kramer, C., Kaspar, K., & König, J. (2024). Measuring pre-service teachers' decision-making in classroom management: A video-based assessment approach. *Teaching and Teacher Education*, 138, Article 104426. <https://doi.org/10.1016/j.tate.2023.104426>
14. Li, L., Hu, Y., Yang, X., Wu, M., Tao, P., Chen, M., & Yang, C. (2024). Enhancing pre-service teachers' classroom management competency in a large class context: The role of fully immersive virtual reality. *Humanities and Social Sciences Communications*, 11, Article 1050, 1–14. <https://doi.org/10.1057/s41599-024-03538-9>

15. Kramer, C., Strauß, S., & König, J. (2025). Student teachers' learning about effective classroom management: Efficacy and transfer of a video-based intervention aimed at promoting situation-specific skills and self-efficacy. Zeitschrift für Erziehungswissenschaft, 28, 1501–1523. <https://doi.org/10.1007/s11618-025-01359-3>

Information about authors:

Kastayev Serik - PhD doctor, Central Asian Innovation University, Shymkent, Kazakhstan, e-mail: kastaev_s@mail.ru

Tursynbayeva Ainur - PhD Doctoral Student, of M. Auezov South Kazakhstan University, Shymkent, Kazakhstan, e-mail: t_ainur81@mail.ru

Sartayeva Nazgul (corresponding author) - PhD, senior lecturer, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan, e-mail: nazgul_88s@mail.ru

Mambetaliyev Seribay - Senior Lecturer, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan, e-mail: seribay.mambetaliyev@ayu.edu.kz

Bostanova Ardak - Candidate of Biological Sciences, Ass. Professor. Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan, e-mail: ardak.bostanova@ayu.edu.kz

Авторлар туралы мәлімет:

Кастаев Серик - PhD доктор, Орталық Азия инновациялық университеті, Шымкент қ., Қазақстан, e-mail: kastaev_s@mail.ru

Турсынбаева Айнур - PhD докторант, М. Әуезов атындағы Оңтүстік Қазақстан зерттеу университеті, Шымкент қ., Қазақстан, e-mail: t_ainur81@mail.ru

Сартаева Назгуль (автор-корреспондент) - PhD, аға оқытушы, Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қаласы, Қазақстан, e-mail: nazgul_88s@mail.ru

Мамбеталиев Серибай - аға оқытушы, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан, Қазақстан, e-mail: seribay.mambetaliyev@ayu.edu.kz

Бостанова Ардак - биология ғылымдарының кандидаты, доцент м.а., Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан, Қазақстан, e-mail: ardak.bostanova@ayu.edu.kz

Сведения об авторах:

Кастаев Серик - PhD доктор, Центрально-Азиатский инновационный университет, Шымкент, Казахстан, e-mail: kastaev_s@mail.ru

Турсынбаева Айнур - PhD докторант, Южно-Казахстанский научно-исследовательский университет им. М.Ауэзова, Шымкент, Казахстан, e-mail: t_ainur81@mail.ru

Сартаева Назгуль (автор-корреспондент) - PhD, старший преподаватель, Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан, e-mail: nazgul_88s@mail.ru

Мамбеталиев Серибай - старший преподаватель, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркестан, Казахстан, e-mail: seribay.mambetaliyev@ayu.edu.kz

Бостанова Ардак - кандидат биологических наук, и.о. доцента, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркестан, Казахстан, e-mail: ardak.bostanova@ayu.edu.kz

У.З. Ешимова^{1*}, Б.Б. Махмұтов², Ж.О. Оразымбетова³, З.Н. Маханбетова⁴

¹Қорқыт ата атындағы Қызылорда Университеті, Қызылорда қ., Қазақстан Республикасы

²ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана қ., Қазақстан Республикасы

³Қызылорда педагогикалық жоғары колледжі, Қызылорда қ., Қазақстан Республикасы

⁴Астана халықаралық университетінің жоғары педагогикалық колледжі, Астана қ., Қазақстан Республикасы

*e-mail: ultai2005@gmail.com

¹ORCID 0009-0005-2832-9793, ²ORCID 0000-0002-1513-7056, ³ORCID 0009-0009-3010-2976

⁴ORCID 0009-0001-6774-0965

ПЕДАГОГИКАЛЫҚ КОЛЛЕДЖ СТУДЕНТТЕРІНІҢ ПЕДАГОГИКАЛЫҚ ПРАКТИКАСЫНДА КӘСІБИ ЖӘНЕ ЖҰМСАҚ DAҒДЫЛАРЫН БАҒАЛАУДЫҢ ПЕДАГОГИКАЛЫҚ ШАРТТАРЫ МЕН ӨЛШЕМДЕРІ

Мақалада педагогикалық колледж студенттерінің педагогикалық практика барысында қалыптасатын кәсіби (hard skills) және жұмсақ (soft skills) дағдыларын бағалаудың педагогикалық шарттары мен өлшемдері қарастырылады. Зерттеу бойынша алынған студенттердің кәсіби құзіреттілігін, коммуникативтік, ұйымдастырушылық, рефлексивтік дағдыларын анықтауға бағытталған бағалау критерийлері мен көрсеткіштері ұсынылады. Зерттеудің мақсаты - студенттердің жас ерекшелігін ескере отырып, осы дағдыларды бағалаудың педагогикалық шарттарын негіздеу, өлшемдері мен көрсеткіштерін нақтылау. Зерттеуде ғылыми әдебиеттерді талдау, сауалнама, өзін-өзі бағалау, өзара бағалау, педагогикалық бақылау және модельдеу әдістері қолданылды. Эмпирикалық деректер Астана және Қызылорда қалаларындағы педагогикалық колледждерде, 01140100 – «Бастауыш білім беру педагогикасы мен әдістемесі» мамандығы бойынша 2024-2026 жылдары жинақталды. Зерттеу барысында педагогикалық практикаға арналған дағдылар картасы әзірленіп, hard skills пен soft skills-ті бағалаудың құрылымдық компоненттері мен критерийлері айқындалды. Практикаға дейінгі және кейінгі өзін-өзі бағалау нәтижелері студенттердің бірқатар дағдылары бойынша оң өзгерісті көрсетті; топтық рефлексия олардың әрі қарай кәсіби дамуына қажет бағыттарды нақтылауға мүмкіндік берді. Ұсынылған өлшемдер педагогикалық практика үдерісінде студенттердің кәсіби және тұлғалық әлеуетін кешенді бағалауды қамтамасыз етеді. Бағалау нәтижелері студенттердің кәсіби даярлық деңгейін арттыруға және педагогикалық қызметке бейімделу үдерісін жетілдіруге негіз болады.

Түйін сөздер: педагогикалық практика, hard skills, soft skills, болашақ мұғалімдер, бағалау, рефлексия, педагогикалық колледж.

Кіріспе

Педагогикалық практика – болашақ мұғалімнің кәсіби даярлығын нақты мектеп жағдайында байқататын негізгі кезеңдердің бірі. Осы уақытта студент сабақ жоспарлайды және өткізеді, оқу жетістіктерін бағалайды, тәрбие жұмысын ұйымдастырады, оқушылармен және педагогтермен қарым-қатынас орнатады. Яғни практикада оқу барысында меңгерілген теориялық білімнің кәсіби әрекетке қалай ауысатыны көрінеді. Сондықтан студенттің даярлығын тек пәндік біліммен немесе сабақ өткізу техникасымен шектеу оның нақты кәсіби мүмкіндігін толық сипаттамайды.

Мұғалімнің күнделікті жұмысы пәндік және әдістемелік біліммен қатар, басқа адамдармен әрекеттесу, жағдайға бейімделу, шешім қабылдау және өз әрекетін талдау қабілеттерін талап етеді. Болашақ педагог үшін бұл дағдылардың маңызы практика кезінде айқын көрінеді: бір сабақтың өзінде студент оқу мақсатын түсіндіреді, сыныптың жұмысын ұйымдастырады, кері байланыс береді және күтпеген жағдаяттарға жауап береді. Осы міндеттер кәсіби hard skills пен тұлғааралық және метапәндік soft skills-тің өзара байланыста қолданылатынын көрсетеді.

Біздің зерттеуде педагогикалық практика осы екі дағды тобын табиғи кәсіби ортада қатар байқауға мүмкіндік беретін кезең ретінде қарастырылды. Студенттің сабақ жоспары, пәндік білімді қолдануы, бағалау құралдарын әзірлеуі немесе цифрлық ресурстарды пайдалануы кәсіби дағдылардың көрінісі болса, оқушымен тіл табысуы, командада жұмыс істеуі, шешім қабылдауы және рефлексия жасауы жұмсақ дағдылардың көрінуімен байланысты.

Мәселе бұл дағдылардың бағалану тәсілдерінде туындайды. Hard skills-ті сабақ жоспары, практикалық тапсырма немесе орындалған жұмыс нәтижесі арқылы салыстырмалы түрде нақты бақылауға болады. Soft skills-ті бағалауда жағдай күрделірек: нәтиже бағалаушының пікіріне, студенттің өзін-өзі бағалау дәлдігіне және нақты әрекеттен жиналған дәлелдердің көлеміне тәуелді болуы мүмкін. Сондықтан педагогикалық практикада бір ғана бағалау тәсіліне сүйенбей, бірнеше дерек көзін қатар қолдану қажет.

Педагогикалық практиканы бағалау осы себептен сабақ жоспарының сапасы немесе өткізілген сабақ саны сияқты көрсеткіштермен ғана шектелмеуі тиіс. Болашақ мұғалімнің кәсіби даярлығы пәндік-әдістемелік әрекетпен бірге коммуникация, ынтымақтастық, сыни ойлау, шығармашылық, көшбасшылық, мәселені шешу, цифрлық сауаттылық және рефлексия сияқты дағдылардан да көрінеді. Зерттеуде бағалау мазмұны осы екі бағытты бір жүйеде қарастыруға негізделді.

Ғылыми еңбектерде soft skills құрамы бірдей берілмейді: зерттеу мақсатына қарай коммуникация, командада жұмыс істеу, сыни ойлау, шығармашылық, эмоционалдық интеллект, көшбасшылық, бейімделу, уақытты басқару және проблемаларды шешу дағдылары әртүрлі үйлесімде қарастырылады. Hard skills көбіне нақты кәсіби қызметті орындауға қажет пәндік, әдістемелік, технологиялық және практикалық әрекеттермен байланыстырылған. Педагог қызметінде бұл екі топты бөлек қарастыру шартты сипатта: мысалы, сабақ мазмұнын дұрыс түсіндіру үшін пәндік білім қажет болса, сол мазмұнды сыныпқа жеткізу үшін коммуникация, ұйымдастыру және рефлексия дағдылары қатар іске қосылады.

Педагогикалық практикадағы дәстүрлі есеп беру көбіне өткізілген сабақтарды, тәрбие жұмыстарын және тәлімгер мен практика жетекшісінің пікірін тіркеуге бағытталады. Бұл материалдар маңызды болғанымен, студенттің дағдыларының қалай қалыптасқанын кезең-кезеңімен көрсетуге әрдайым жеткіліксіз. Сондықтан зерттеуде практика мазмұнын құзыреттілік тұрғысынан қарастырып, студент орындайтын әрекеттерді hard skills және soft skills көрсеткіштерімен байланыстыру ұсынылды.

Зерттеуде кәсіби және жұмсақ дағдылардың жұмысқа алынған құрамы 1-кестеде берілді.

Кесте 1 – Hard skills және soft skills - дағдылары

Hard skills	Soft skills
• сабақ жоспарлау; • пәндік білімді қолдану; • тәжірибе қою; • цифрлық құралдарды пайдалану; • бағалау құралдарын әзірлеу; • зерттеу жүргізу.	• коммуникация; • ынтымақтастық; • көшбасшылық; • сыни ойлау; • креативтілік; • уақытты басқару; • рефлексия.

Практика кезінде бұл дағдыларды бағалау студенттің тек не білетінін емес, сол білімді нақты кәсіби міндетте қалай қолданатынын көруге мүмкіндік береді. Осыған байланысты зерттеуде бір реттік бағалаудың орнына бірнеше тәсілді үйлестіру қарастырылды: практикаға дейінгі және кейінгі өзін-өзі бағалау, өзара және сараптамалық бағалау, практикалық әрекетті бақылау, портфолио, жағдаяттық тапсырмалар, жеке және топтық

рефлексия. Мұндай деректерді қатар қарастыру әр тәсілдің жеке шектеулерін азайтып, студенттің кәсіби дамуындағы өзгерісті толығырақ сипаттауға мүмкіндік береді.

Осы жұмыстың міндеті – педагогикалық практика жағдайында hard skills пен soft skills-ті қандай шарттарда және қандай белгілер арқылы бағалауға болатынын нақтылау.

Мақалада педагогикалық шарттар бағалауды ұйымдастыру ортасын, өлшемдер дағдының қалыптасу сапасын, ал көрсеткіштер студент әрекетінен тікелей байқалатын белгілерді сипаттайтын ұғымдар ретінде қолданылды. Осындай жіктеу бағалаушының жалпы әсеріне сүйенуді азайтып, бақылауды нақты әрекеттерге бағыттауға көмектеседі.

Зерттеудің өзектілігі де осыдан туындайды: педагогикалық практиканың нәтижесін тек орындалған жұмыс көлемімен емес, студенттің кәсіби және жұмсақ дағдыларының практика барысында қалай көрінетіні және өзгертетіні арқылы бағалау қажет.

Зерттеудің мақсаты: педагогикалық колледж студенттерінің жас ерекшелігін ескере отырып, педагогикалық практика кезінде олардың кәсіби және жұмсақ дағдыларын бағалаудың педагогикалық шарттарын теориялық тұрғыдан негіздеу және оларды бағалаудың өлшемдері мен көрсеткіштерін әзірлеу.

Зерттеу болжамы: педагогикалық практика мазмұны кезең-кезеңімен ұйымдастырылып, кәсіби және жұмсақ дағдыларды қатар дамытуға бағытталған модель, жүйелі рефлексия және тәлімгерлік қолдау қолданылған жағдайда студенттердің кәсіби құзыреттілігі, педагогикалық бейімделуі, коммуникациялық мәдениеті және практикалық әрекет сапасы жақсарады.

Зерттеу нысаны - педагогикалық колледж студенттерінің (болашақ мұғалімдердің) педагогикалық практика үдерісі.

Зерттеу пәні - педагогикалық практика барысында болашақ мұғалімдердің soft skills және hard skills дағдыларын бағалау үдерісі.

Педагогикалық колледждердегі практика ұзақтығы курсқа қарай өзгереді: 2-курста – 6 апта, 3-курста – 10 апта, 4-курста – 22 апта. Алайда мектептегі практиканың жекелеген кезеңдері 2 немесе 4 апталық блоктармен өтуі мүмкін. Практика базасының және сыныптың ауысуы студентке қысқа мерзім ішінде жаңа оқушылармен байланыс орнатуды, мектептің жұмыс тәртібіне бейімделуді және кәсіби міндеттерін қатар орындауды талап етеді. Осындай жағдайда сабақтың әдістемелік жағына басымдық беріліп, коммуникация, көшбасшылық, эмоционалдық тұрақтылық, сыни ойлау, командалық жұмыс және рефлексия сияқты soft skills-ті мақсатты дамытуға уақыт жеткіліксіз болуы мүмкін.

Практиканың мазмұны да курстар бойынша бірдей емес. 2-курс студенті көбіне бақылаушы және көмекші рөлінде әрекет етсе, 3-4-курстарда студент дербес сабақ өткізіп, сынып жетекшісінің жұмысына қатысады. Осы айырмашылықтарды ескермей барлық курсқа бірдей талап қою дағдылардың қалыптасу динамикасын дәл көрсете бермейді. Сондықтан модельде hard skills пен soft skills-ті практика кезеңіне сай біртіндеп күрделендіру қағидаты қолданылды.

Осы тәсіл педагогикалық практиканы тек сабақ өткізу кезеңі емес, студенттің пәндік-әдістемелік және тұлғааралық дағдыларын қатар байқап, дамытуға болатын кәсіби орта ретінде қарастыруға негіз болды. Зерттеудің практикалық мәні – практика кезеңдеріне сай дағдылар картасын құру және оны нақты бағалау көрсеткіштерімен байланыстыру.

Зерттеу материалдары мен әдістемесі

Зерттеудің теориялық негізін нақтылау үшін болашақ мұғалімдердің soft skills дағдыларын қалыптастыру мен педагогикалық практикада бағалауға арналған еңбектер қарастырылды. R. Ağcam және A. Doğan (2021) [1] Түркияның мемлекеттік университеттеріндегі болашақ мұғалімдердің коммуникация, проблеманы шешу, уақытты басқару, командалық жұмыс және көшбасшылық дағдыларын зерттеп, нәтижелердің мамандық пен оқу жылына байланысты өзгертетінін көрсеткен. C. Amorim және E. Ribeiro-Silva (2025) [2] дене шынықтыру мұғалімдерін даярлау контекстінде педагогикалық практика

кезінде студенттердің өзіне сенімділік, коммуникация, эмпатия, көшбасшылық және командалық жұмыс дағдыларындағы өзгерістерді сипаттаған.

S. M. Macqual, U. K. M. Salleh және H. Zulnaidi (2021) [3] 722 болашақ мұғалімнің деректерін PLS-SEM әдісімен талдап, soft skills, сыныпты басқару және педагогикалық практика нәтижелері арасындағы байланысты қарастырған. Авторлар soft skills-ті мұғалім даярлау бағдарламасына мақсатты түрде енгізіп, оны бағалауды ұсынады. S. Mentari, U. Nuraini, R. Palil және V. Kitnuntaviwat (2026) [4] студент-мұғалімдердің soft skills дағдыларын бағалауда бірнеше бағалау көзін қолданған. Бұл тәсіл біздің зерттеуде өзін-өзі бағалау, өзара бағалау және педагогикалық бақылауды қатар пайдалану логикасымен үндеседі.

J. Heckman және T. Kautz (2012) [5] стандартты академиялық тестілер мотивация, тұлғалық қасиеттер және әлеуметтік-психологиялық сипаттарды толық қамтымайтынын көрсетеді. Бұл тұжырым soft skills-ті тек тест нәтижесімен бағалаудың жеткіліксіз екенін түсіндіруге маңызды. Сол себепті зерттеуде hard skills үшін орындалған кәсіби әрекеттің нәтижесі, ал soft skills үшін мінез-құлықтың нақты көріністері мен бірнеше бағалау көздері қатар қарастырылды.

Педагогикалық практикаға арналған бағалау құралдарын салыстыру тұрғысынан L. Rusznyak және L. Österling (2024) [6] еңбегі ерекше назарға алынды. Авторлар Оңтүстік Африка, Үндістан, Англия, Сингапур, Канада және Швецияда қолданылатын 10 бағалау рубрикасын салыстырған. Бұл зерттеу педагогикалық практиканы бағалауда тек сабақ өткізу нәтижесін емес, мұғалім әрекетінің бірнеше компонентін бір мезгілде ескеру қажеттігін көрсетеді.

Қазақстандық зерттеулерде де болашақ педагогтердің кәсіби және жұмсақ дағдылары қарастырылған. Л.А. Шкутина, Т.Г. Искакбаева, Ж.А. Карманова және Я. Данек (2022) [7] цифрландыру жағдайында hard skills-ті кәсіби білім мен практикалық әрекеттермен, ал soft skills-ті икемділік, шығармашылық және тұлғааралық сипаттармен байланыстырады. Авторлар екі дағды тобының да болашақ педагогтің кәсіби ортаға бейімделуіндегі маңызын атап көрсетеді.

Қазақстандық еңбектердің қатарында Ж.А. Есказинова, Л.С. Сырымбетова, А.К. Шаймерденова (2025) [8], сондай-ақ С. Жунусова, Л. Байманова, Л. Навиу (2026) [9] зерттеулері аталады. Бұл жұмыстарда болашақ мұғалімдердің soft skills дағдыларын қалыптастыруға арналған құрылымдық-функционалдық тәсілдер қарастырылып, когнитивтік, әлеуметтік-коммуникативтік және өзін-өзі реттеу компоненттері өзара байланыста сипатталады.

ЭЫДҰ-ның «Learning Compass 2030» (2019) [10] тұжырымдамасында білім алушының болашаққа дайындығы білім, дағды, қатынас және құндылықтардың өзара байланысы арқылы қарастырылады. Құжатта сыни ойлау, проблемаларды шешу, ынтымақтастық, өзін-өзі реттеу және бейімделгіштік өзгермелі ортада әрекет етуге қажет дағдылар ретінде көрсетілген.

Қарастырылған еңбектердің басым бөлігі университет студенттеріне арналған. Ал осы зерттеудің нысаны – педагогикалық колледж студенттері. Колледж бен университет студенттерінің жас кезеңі, оқу тәжірибесі және кәсіби өзін-өзі анықтау деңгейі бірдей болмауы мүмкін. Сондықтан бағалау құралдарын таңдағанда осы айырмашылықтарды ескеру қажет деп қабылданды. [11–14] еңбектеріне сүйене отырып, зерттеуде танымдық, физиологиялық және психологиялық-тұлғалық сипаттамалар үш бағыт бойынша жүйеленіп, 2-кестеде берілді.

Кесте 2 - Университет және колледж студенттерінің танымдық, психологиялық физиологиялық ерекшеліктерінің сипаттамалары

1. Танымдық (когнитивті) ерекшеліктері

Ойлау стилі мен әрекеті:	
<i>Колледж студенттері (15-18 жас):</i> Көбіне нақты-бейнелі және практикалық ойлауға бейім. Оларға теориялық ұғымдардан гөрі «мұны іс жүзінде қалай жасау керек?» деген нақты алгоритмдер мен үлгілер маңыздырақ. Ақпаратты қабылдағанда визуалды және еліктеушілік механизмі басым болады.	<i>Университет студенттері (18-23 жас және жоғары):</i> Абстрактілі-теориялық және жүйелі ойлау қабілеті толық қалыптасқан. Олар құбылыстардың себеп-салдарын терең талдап, ғылыми болжамдар жасауға, сын тұрғысынан ойлап, күрделі теориялық концепцияларды оқу үдерісіне интеграциялауға қабілетті.
Ақпаратты меңгеру және зейін	
<i>Колледж:</i> Зейіннің тұрақтылығы кейде эмоция мен қызығушылыққа тәуелді болады. Ақпаратты қабылдауда қысқа мерзімді және оперативті жады белсенді жұмыс істейді.	<i>Университет:</i> Күрделі, көп деңгейлі ақпараттық ағындармен жұмыс істеуге, академиялық мәтіндерді өз бетінше өңдеуге және ұзақ мерзімді жүйелі жадыны тиімді қолдануға дағдыланған

2. Физиологиялық және нейробиологиялық ерекшеліктері

Мидың префронтальды қыртысының дамуы:	
<i>Колледж:</i> Префронтальды қыртыс (мидың жоғары деңгейдегі бақылау, жоспарлау, импульстерді тежеу және жауапкершілік үшін жауапты бөлігі) әлі де қалыптасу үдерісінде. Сондықтан жасөспірімдік максимализм, эмоцияның импульсивті басым болуы, тәуекелге бел буу немесе жағдайды уақытша бағалау жиі кездеседі.	<i>Университет:</i> Префронтальды қыртыс (мидың жоғары деңгейдегі бақылау, жоспарлау, импульстерді тежеу және жауапкершілік үшін жауапты бөлігі) қалыптасқан. Сондықтан, өзін өзі саналы басқару, жағдайды нақты бағалай алады.
Биологиялық ырғақтар мен энергетикалық әлеует:	
<i>Колледж:</i> Ағзаның қарқынды гормональды қайта құрылу кезеңі аяқталып жатқандықтан, физиологиялық энергия мен сергектік жоғары, бірақ эмоционалды және физикалық шаршау (лабильділік) тез байқалуы мүмкін.	<i>Университет:</i> Физиологиялық тұрғыдан ағза тұрақтанған. Энергия қоры мен жұмысқа қабілеттілік ұзақ мерзімді жүктемелерге (мысалы, сессия кезіндегі ауыр жұмыс) төтеп беруге бейімделген.

3. Психологиялық және тұлғалық ерекшеліктері

Кәсіби сәйкестік және өзін-өзі анықтау:	
<i>Колледж:</i> «Мен кімін?» және «Мамандықты дұрыс таңдадым ба?» деген экзистенциалды сұрақтар осы кезеңде өте өткір тұрады. Олар көбінесе мектептен кейін бірден келетіндіктен, балалық шақ пен ересек өмірдің арасындағы өтпелі кезеңді бастан кешіреді. Педагог бейнесін көбінесе колледж оқытушыларының немесе сүйікті мектеп мұғалімдерінің көшірмесі ретінде қабылдайды.	<i>Университет:</i> Кәсіби өзін-өзі санасы әлдеқайда тұрақты. Олар өз мамандығын саналы түрде таңдаған (немесе бұл бағытта ересек тәжірибеден өткен). Олардың психологиялық портреті дербестікке, автономияға және авторлық ұстанымға негізделген.
Қарым-қатынас және әлеуметтену:	
<i>Колледж:</i> Құрдастар тобындағы (референтті топ) пікір маңызды рөл атқарады. Әлеуметтік бейімделу үдерісі қарқынды жүреді.	<i>Университет:</i> Жеке тұлғалық шекаралар анық, кәсіби ортадағы (әріптестермен, мектеп басшылығымен) қарым-қатынаста іскерлік этикет пен серіктестік басым болады.

2-кестеде көрсетілген ерекшеліктер практикаға қатысатын студенттерді жеке-жеке жіктеу үшін емес, оқытуды және тәлімгерлік қолдауды жоспарлау үшін пайдаланылды. 2-курс студенттеріне нақты нұсқаулық, үлгі және қадамдық қолдау көбірек қажет болуы

мүмкін; ал тәжірибесі артқан студенттерге дербес шешім қабылдауға көбірек мүмкіндік беріледі. Зерттеу 2-курста педагогикалық практика басталатын кезеңге ерекше назар аударып, кәсіби біліммен қатар жұмсақ дағдылардың бастапқы көріністерін де бақылауды көздеді. Осы мақсатта практикаға дейін және екі апталық практикадан кейін «Өрмекші торы» әдісімен өзін-өзі бағалау жүргізілді. Оған Астана және Қызылорда қалаларындағы педагогикалық колледж студенттері қатысты.

Практика басталғанға дейін студенттерге бағдарлау семинары өткізілді: тәлімгерлер, практика базалары және күнделік мазмұны түсіндірілді. Зерттеу аясында осы қалыпты рәсімге қосымша ретінде әр студентке практика мерзіміне арналған дағдылар картасы берілді. Семинарда картаны толтыру тәртібі мен туындаған сұрақтар талқыланды. 9-сыныптан кейін педагогикалық мамандықты таңдаған студенттерге оқушылармен қарым-қатынас орнату, олардың қызығушылығын қолдау және ойын элементтерін орынды қолдану мәселелеріне ерекше назар аударылды.

Кесте 3 - Дағдылар картасы

Бастауыш сынып мұғалімдерінің педагогикалық практикасы (2 апталық кезең)

Практика мақсаты	Қалыптастырылатын жұмсақ дағдылар	Қалыптастырылатын кәсіби дағдылар	Қалыптасу көрсеткіштері
Мектеп ортасымен танысу	Бейімделу, жауапкершілік, әдепті қарым-қатынас	Мектептің нормативтік құжаттарымен танысу	Ішкі тәртіп ережелерін сақтайды, құжаттарды біледі
Сабақтарды бақылау	Бақылау, аналитикалық ойлау, рефлексия	Сабақ құрылымын талдау, оқыту әдістерін анықтау	Сабаққа талдау жасайды, тиімді әдістерді ажыратады
Оқушылармен қарым-қатынас орнату	Коммуникация, эмпатия, тыңдай білу	Жас ерекшеліктерін ескере отырып педагогикалық қарым-қатынас құру	Оқушылармен сенімді байланыс орнатады
Мұғалімнің жұмысына көмек көрсету	Ынтымақтастық, командада жұмыс, бастамашылдық	Оқу материалдарын әзірлеуге қатысу, сынып жұмыстарын ұйымдастыру	Дидактикалық материалдар дайындайды
Сабақтың жекелеген кезеңдерін өткізу	Өзін-өзі басқару, сенімділік, уақытты басқару	Сабақтың кіріспесі, сергіту сәті, қорытынды кезеңдерін өткізу	Сабақтың жеке элементтерін тиімді ұйымдастырады
Практиканы қорытындылау	Рефлексия, өзін-өзі бағалау	Практика нәтижесін талдау, есеп дайындау	Рефлексивтік есеп жазады, кәсіби даму бағыттарын анықтайды

Кесте 4 - Дағдылар картасы

Бастауыш сынып мұғалімдерінің педагогикалық практикасы (4 апталық кезең)

Практика мақсаты	Қалыптастырылатын жұмсақ дағдылар	Қалыптастырылатын кәсіби дағдылар	Қалыптасу көрсеткіштері
Оқу-тәрбие процесін дербес ұйымдастыру	Көшбасшылық, жауапкершілік, бастамашылдық	Қысқа мерзімді жоспар әзірлеу және сабақ өткізу	Сабақ жоспарын талапқа сай құрастырады және өткізеді

Белсенді оқыту әдістерін қолдану	Шығармашылық, икемділік, мәселе шешу	Саралап оқыту, белсенді әдістер, цифрлық ресурстарды пайдалану	Өртүрлі оқыту әдістерін тиімді қолданады
Бағалау үдерісін жүзеге асыру	Әділдік, объективтілік, кері байланыс беру	Қалыптастырушы бағалау, дескрипторлар әзірлеу, кері байланыс ұйымдастыру	Бағалау құралдарын дұрыс қолданады
Инклюзивті білім беру қағидаларын сақтау	Толеранттылық, эмпатия	Оқушылардың жеке ерекшеліктерін ескере отырып оқыту	Сараланған тапсырмалар әзірлейді
Ата-аналармен және педагогтермен ынтымақтастық	Іскерлік қарым-қатынас, келіссөз жүргізу, командалық жұмыс	Ата-аналармен байланыс орнату, әдістемелік бірлестік жұмысына қатысу	Кәсіби қарым-қатынас мәдениетін көрсетеді
Сыныпты басқару	Эмоционалдық тұрақтылық, шешім қабылдау	Тәртіпті ұйымдастыру, оқу ортасын тиімді басқару	Сыныптағы оқу процесін тиімді ұйымдастырады
Зерттеушілік қызмет	Сыни ойлау, деректерді талдау	Сабақты зерттеу элементтері, оқыту нәтижелерін талдау	Сабақ нәтижелерін талдап, қорытынды жасайды
Кәсіби даму	Өзін-өзі дамыту, рефлексия	Портфолио қалыптастыру, кәсіби жетістіктерді жинақтау	Практика бойынша портфолио мен қорытынды есеп әзірлейді

2025–2026 оқу жылында дағдылар картасы студенттерге алғаш рет ұсынылды; оны 2026–2027 оқу жылында жалғастыру жоспарланған. Өзара бағалауды ұйымдастыру үшін Е.А. Паньконың «Мұғалімнің педагогикалық бағыттылығын анықтау» [15] тестінің логикасы зерттеу мақсатына бейімделді. Түпнұсқа құрал студенттің кәсіби-тұлғалық бағытын сипаттауға арналғандықтан, оның жекелеген жағдаяттары hard skills пен soft skills көріністерін ажыратуға икемделді. Бейімдеу барысында тармақтар екі бағытқа бөлінді:

✓ Hard skills-ке бағытталған тармақтар ретінде - пәндік білімді беру, сабақ жоспарын жүйелеу, әдістемелік құралдарды, цифрлық технологияларды қолдану және ақпаратты дәл жеткізуге бағытталған әрекеттерді бағалау;

✓ Soft skills-ке бағытталған тармақтар ретінде - оқушылармен қарым қатынас (коммуникация), эмоционалдық қолдау, сыныптағы психологиялық жағдайды басқару, стресске төзімділік және топтағы өзара іс-қимылды ұйымдастыруды бағалауды қамтиды. Сабақтан соң, студенттер тәлімгермен бірлесе өзара диалог ұйымдастырып сабақ үдерісін талқылау жүргізеді. Тәлімгер студенттерің өзара бір-бірін бағалауын ұйымдастыруды Е.А. Панько тестінің бейімделген нұсқасы арқылы жүргізуді қолданды. Нұсқа үлгісі мен орындау тәртібі төменде беріледі.

Төменде Сіздердің педагогикалық практикаларыңызға қатысты бірнеше жағдаяттар берілген. Әрбір жағдаят бойынша Сізге екі нұсқаның (А немесе Б) ішінен өзіңізге ең жақын, көбірек қолданатын немесе маңызды деп санайтын әрекетті таңдау қажет.

1. Сабаққа дайындық барысында Сіз үшін не маңыздырақ?

А (Hard): Тақырыпты академиялық тұрғыдан терең зерттеп, көрнекіліктер, кестелер және цифрлық ресурстарды (презентация, платформалар) мінсіз дайындау.

Б (Soft): Оқушылардың көңіл-күйін ескеріп, сабақты қалай қызықты, эмоциялық тұрғыдан тартымды және еркін ортада өткізуді ойластыру.

2. Сабақ үстінде түсіндіріп жатқан материалыңызды оқушылар түсінбей қалғанын байқасаңыз, не істейсіз?

А (Hard): Басқа әдістемелік тәсілді, формуланы немесе алгоритмді қолданып, материалды ғылыми тұрғыдан қайтадан дұрыс түсіндіруге тырысамын.

Б (Soft): Оқушылардың қобалжуын немесе сенімсіздігін сейілту үшін мысалдар келтіріп, қолдау сөздер айтып, жағдайды жұмсартамын.

3. Сіздің ойыңызша, мұғалімнің ең басты жетістігі неде?

А (Hard): Оқушылардың пән бойынша терең білім алып, тесттер мен емтихандардан жоғары нәтиже көрсетуі.

Б (Soft): Оқушылардың өзіне деген сенімділігінің артуы, мұғаліммен және бір-бірімен жақсы қарым-қатынас орната алуы.

4. Күтпеген қиындық немесе сыныпта тәртіп бұзылған жағдайда Сіздің бірінші әрекетіңіз қандай болады?

А (Hard): Уақытты жоғалтпай, нақты жоспарланған регламент пен тәртіп ережелерін қолданып, жұмысты жалғастырамын.

Б (Soft): Сабыр сақтап, эмоционалды интеллектіні қолдана отырып, оқушылармен жеке тіл табысу арқылы мәселені шешуге тырысамын.

5. Сабақ соңында кері байланыс алған кезде Сіз нені көбірек қалайсыз?

А (Hard): Әдіскердің немесе тәлімгердің сабақтың әдістемелік құрылымы, мазмұны бойынша нақты кәсіби ескертпелерін тыңдау.

Б (Soft): Оқушылардың сабақ ұнады ма, жоқ па, көңіл-күйлері қандай болғаны туралы шынайы пікірін білу.

Бейімделген Е.А. Панько тесті өзара бағалауда студенттің сабағына жалпы «жақсы/нашар» деген баға берудің орнына нақты әрекеттерді талқылауға мүмкіндік берді. Сабақтан кейін студент, топтасы және тәлімгер қандай жағдайда қандай әрекет таңдалғанын талқылады. Осы тәсіл арқылы кері байланыс әдістемелік шешімдер мен қарым-қатынас ерекшеліктеріне сүйеніп берілді. Өзара бағалауда байқалған негізгі мүмкіндіктер төмендегідей болды:

1. Студент өз әрекетін басқа адамның бақылауы арқылы салыстыра алады. Практикант өзінің сабағын бір қырынан қабылдауы мүмкін, ал бақылаушы оның оқушылармен эмоционалдық байланысқа немесе мазмұнды әдістемелік тұрғыдан жеткізуге қай деңгейде басымдық бергенін нақты жағдаяттар арқылы көрсетеді. Бұл кейінгі топтық рефлексияға бастапқы материал береді.

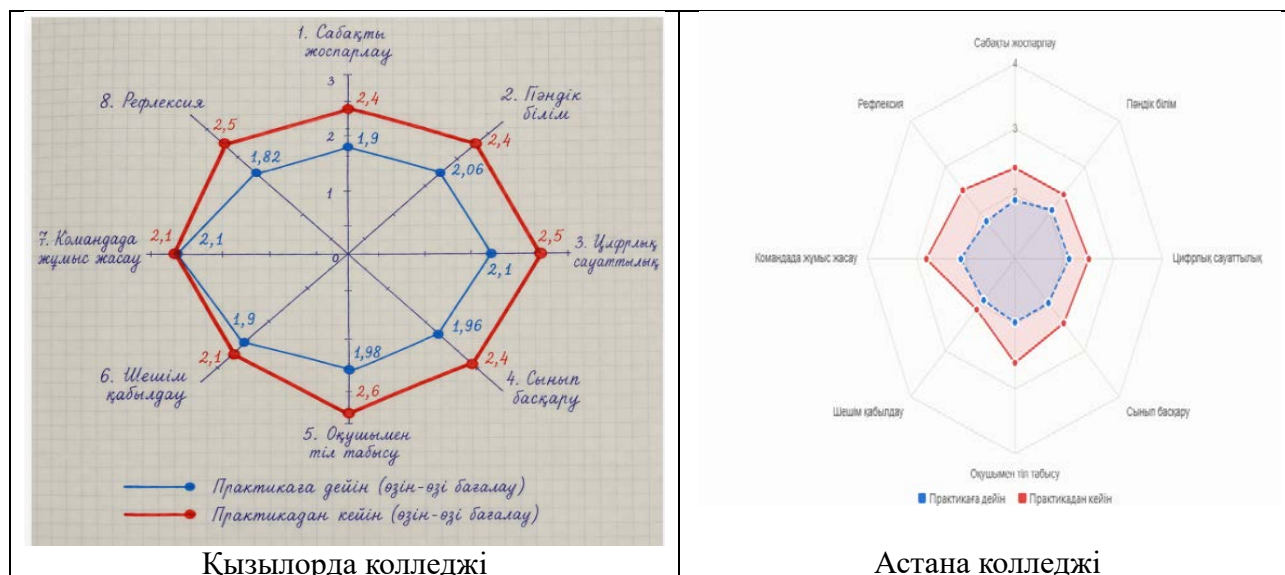
2. Кері байланыс нақты белгілерге сүйенеді. «Сабақ ұнады/ұнамады» деген жалпы пікірдің орнына бағалаушы сабақтағы түсіндіру, кері байланыс, оқушының көңіл күйіне назар аудару немесе сыныпты ұйымдастыру сияқты әрекеттерді атап көрсете алады. Мұндай формат студентке қай дағдысы басым, қай тұсын дамыту қажет екенін түсінуге көмектеседі.

3. Өзара бағалау рефлексияны күшейтеді. Басқа студенттің сабағын белгіленген жағдаяттар бойынша талдау кезінде бақылаушы да өзінің осындай жағдайда қалай әрекет ететінін салыстырады. Осылайша өзара бағалау тек пікір білдіру емес, өзінің кәсіби әрекетін қайта қарауға мүмкіндік беретін рефлексивтік жұмысқа айналады.

4. Талқылау топтағы әртүрлі күшті жақтарды көруге мүмкіндік береді. Бір студент коммуникацияда сенімді болуы мүмкін, екіншісі әдістемелік ұйымдастыруда ерекшеленеді. Тест сұрақтарына сүйенген талқылау осы айырмашылықтарды тұлғалық бағаға айналдырмай, кәсіби әрекет деңгейінде қарастыруға жағдай жасады.

Практика аяқталғаннан кейін жеке есептер дайындалып, одан соң топтық рефлексия «Лотос гүлі» әдісімен өткізілді. Әртүрлі мектептерде практикадан өткен студенттер шағын топтарға біріктіріліп, кездескен қиындықтары мен сәтті шешімдерін ортақ талқылауға шығарды. Әр топ негізгі проблемаларды сегіз бағытқа жинақтап, оларды «Лотос гүлі» сызбасына орналастырды. Әдісті қолдану реті төменде берілген.

1-қадам. Орталық тақырып. Парақтың ортасына талқыланатын негізгі ұғым – «педагогикалық практика» орналастырылады.

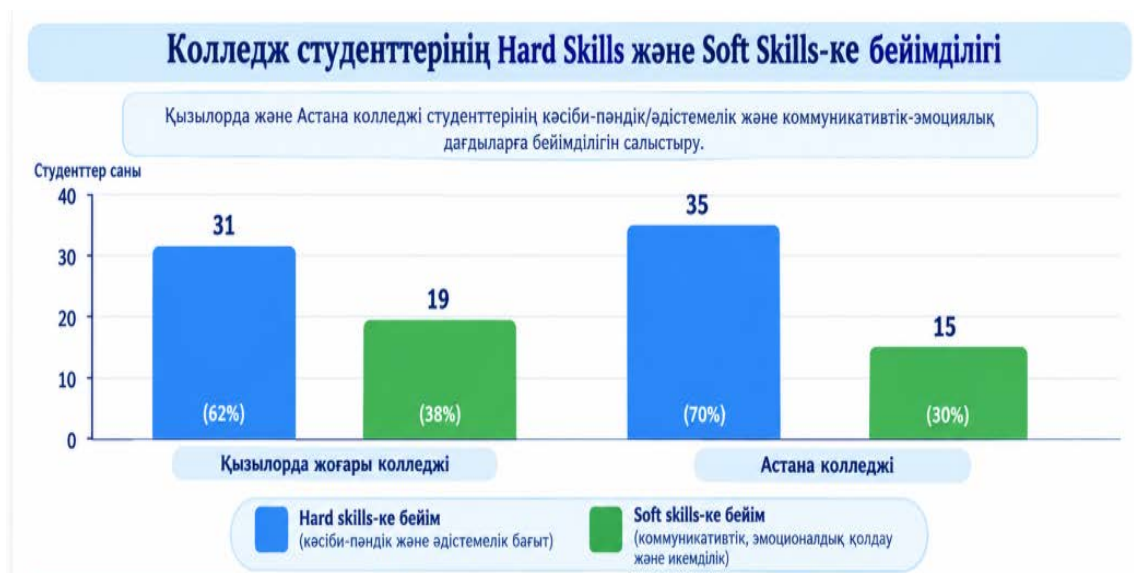


Сурет 3 - Қызылорда және Астана жоғары педагогикалық колледж студенттерінің өзін-өзі бағалауы

Өзін-өзі бағалау нәтижелерінде екі колледж студенттеріне ортақ бірнеше белгі байқалды. Шешім қабылдау дағдысында қиындықтар көрсетіліп, тәлімгер мен сынып жетекшісінің пікіріне сүйену жиі байқалды. Рефлексия мен өзінің практикалық әрекетін сыни талдау дағдыларын дамыту қажеттілігі екі топта да көрінді. Цифрлық сауаттылық бойынша Астана колледжі студенттерінің көрсеткіші сәл жоғары болды. Өзара бағалау нәтижелері 5-кестеде берілген.

Кесте 5 - Өзара бағалау барысында Е.А. Панько тестінің нәтижесі

Оқу орындары	Hard skills-ке бейім	Soft skills-ке бейім
Қызылорда жоғары колледжі студенттері	31 (кәсіби-пәндік және әдістемелік бағытқа) көбірек	19 (коммуникативтілікке, эмоционалдық қолдауға және икемділікке)
Астана колледжінің студенттері	35 (кәсіби-пәндік және әдістемелік бағытқа) көбірек	15 (коммуникативтілікке, эмоционалдық қолдауға және икемділікке)



Сурет 4 – Колледж студенттерін Hard Skills және Soft Skills-ке бейімділік көрсеткіші

Әр колледжден 50 студенттің нәтижесі қарастырылды. Қызылорда жоғары колледжінде 31 студент (62%) кәсіби-пәндік және әдістемелік бағыттағы Hard Skills-ке, 19 студент (38%) коммуникация, эмоционалдық қолдау және икемділікпен байланысты Soft Skills-ке көбірек бейім болды.

Астана колледжінде 35 студент (70%) Hard Skills бағытын, 15 студент (30%) Soft Skills бағытын басым таңдады. Осылайша екі колледжде де кәсіби-пәндік және әдістемелік бағыттағы таңдау үлесі жоғары болды (6-кесте).

Кесте 6 - Колледждердің айырмашылықтары

Көрсеткіш	Қызылорда	Астана	Айырмашылықтары
Hard Skills	31 (62%) - пәндік және әдістемелік жұмыстарға басымдылық береді	35 (70%) - пәндік және әдістемелік жұмыстарға басымдылық береді	Екі колледж студенттерінде де кәсіби дағдылары жоғары екені анықталды.
Soft Skills	19 (38%) - оқушылармен жақын қарым-қатынас орнатуға және атмосфераға көбірек көңіл бөледі	15 (30%) - оқушылармен жақын қарым-қатынас орнатуға және атмосфераға көбірек көңіл бөледі	

Екі топты салыстырғанда Астана колледжінде Hard Skills бағытының үлесі Қызылорда жоғары колледжіне қарағанда 8 пайыздық тармаққа жоғары, ал Soft Skills үлесі 8 пайыздық тармаққа төмен болды. Бұл нәтиже студенттердің пәндік-әдістемелік әрекетке көбірек бағдарланғанын көрсетеді, бірақ педагогикалық практика коммуникация, эмоционалдық қолдау және икемділік сияқты дағдыларды да қатар талап етеді. Сондықтан практика алдындағы даярлықта осы дағдыларды арнайы жағдаяттар, кері байланыс және рефлексия арқылы дамытуға назар аудару қажет.

Практика аяқталғаннан кейін студенттер жеке есептерін дайындап, одан соң «Лотос гүлі» әдісімен топтық рефлексия жүргізді. Талқылау нәтижесінде «Рефлексия», «Шешім қабылдау» және «Басқару» бағыттарында практикаға дейін қосымша даярлық қажет екені белгіленді (5-сурет).

Үмж қарау	Ресурстар жинақтау	Күтілетін нәтижені анықтау	Сабақтан кейін талдау	Қателерді табу	Басқалардың сабағына қатысу	Назар аудару	Тыңдау	Қызықтыру
Мақсатты айқындау	Жоспарлау	Уақытты тиімді пайдалану		Рефлексия		Мадақтау	Оқушымен тіл табысу	Сыйлық беру
Оқу бағдарла масын алу	Бағалау дұрыс	Тапсырма әзірлеу				Сенім білдіру	Жақсы баға қою	Бірге ойнау
АКТ меңгеру	ЖИ қолдану	Интерактив тақтаны қолдану	Жоспарлау	Рефлексия	Оқушымен тіл табысу	Мотивация беру	Барлығына тең қарау	Болған жағдайды жылдам бағалау
Түрлі бағдарламаны қолдану	Цифрлық сауаттылық	Интернеттен сауатты ақпарат алу	Цифрлық сауаттылық	Педагогикалық практика	Басқару	Әділ шешім ұсыну	Басқару	Сенім білдіру
Қауіпсіздікті сақтау	Жаңалықтарды сабақта пайдалану	Бір бірімізге үйрету	Шешім қабылдау	Пәндік білім	Командамен жұмыс	Қадағалау		
Сынып жетекшімен ақылдасу	Тәлімгерден кеңес алу	Шешімді өзгертпеу	Әр пәнге қызықтыру	Сабаққа дұрыс дайындалу	Теория мен практиканы байланыстыру	Ортақ мақсат қою	Бір бірімізді тыңдау	Бір біріне қолдау жасау
	Шешім қабылдау	Дәлел жинау	Бір бірімізге көмектесу	Пәндік білім	Зерттеу жүргізу	Сынып жетекшісімен ақылдасу	Командамен жұмыс	Кері байланысты дұрыс қабылдау
			Білмегенді сұрау кеңес алу	Жаңа әдістерді пайдалану	Әр пәнді дұрыс білу	Сенімді бұзбау	Жетістіктерді бірге бөлісу	Жігі ақылдасу

Сурет 5 - Студенттердің практикадан кейінгі топтық рефлексиясының қорытындысы

Қорытынды

Зерттеу педагогикалық колледж студенттерінің кәсіби даярлығын hard skills және soft skills көрсеткіштерін бірге қарастырғанда анағұрлым толық сипаттауға болатынын көрсетті. Өзара бағалау нәтижесінде Қызылорда жоғары колледжі студенттерінің 62%-ы, Астана колледжі студенттерінің 70%-ы кәсіби-пәндік және әдістемелік бағыттағы Hard Skills-ті басым таңдады. Soft Skills бағыты тиісінше 38% және 30% болды. Бұл қатынас практика алдындағы даярлықта коммуникация, эмоционалдық қолдау, икемділік және тұлғааралық әрекетке арнайы орын беру қажеттігін көрсетеді.

Практикаға дейінгі және кейінгі өзін-өзі бағалау дағдылардың бірдей қарқында өзгермегенін көрсетті. Практикаға дейін сабақ жоспарлау, пәндік білім, цифрлық сауаттылық, сыныпты басқару, оқушымен тіл табысу, шешім қабылдау, командада жұмыс жасау және рефлексия бойынша орташа көрсеткіштер тиісінше 1,90; 2,06; 2,10; 1,96; 1,98; 1,90; 2,10 және 1,82 балл болды. Практикадан кейін олар 2,40; 2,40; 2,50; 2,40; 2,60; 2,10; 2,10 және 2,50 баллға жетті.

Ең үлкен өсім рефлексияда (+0,68 балл) және оқушымен тіл табысуда (+0,62 балл) байқалды. Сабақ жоспарлау +0,50 баллға, сыныпты басқару +0,44 баллға, цифрлық сауаттылық +0,40 баллға өсті. Бұл деректер практика кезінде студенттің тек теориялық білімін қолдану емес, оқушымен қарым-қатынас орнату, жағдайға бейімделу және өзінің әрекетін талдау тәжірибесі де өзгертінін көрсетеді.

Зерттеу барысында осы өзгерістерді жүйелі бақылау үшін дағдылар картасы әзірленді. Онда практика кезеңінің мақсаты, қалыптастырылатын жұмсақ және кәсіби дағдылар, сондай-ақ олардың нақты көрсеткіштері өзара байланыстырылды. Бағалау жүйесінде өзін-өзі бағалау, практикалық әрекетті бақылау және рефлексия нәтижелері бір-бірін толықтыратын дерек ретінде қолданылды.

Алынған нәтижелер педагогикалық практиканың hard skills пен soft skills-ті қатар дамытуға мүмкіндік беретін орта екенін көрсетті. Практиканың нәтижелілігін бағалауда студенттің пәндік және әдістемелік даярлығымен бірге коммуникациясы, әлеуметтік әрекеті, рефлексиясы және шешім қабылдауы да ескерілуі тиіс.

Ұсынылған дағдылар картасы мен топтық рефлексия құралы студенттің әрекетін бір реттік қорытынды баға арқылы емес, практика барысындағы нақты көріністер арқылы талдауға мүмкіндік береді. Осы ақпарат студенттің қай дағдысы жақсы қалыптасқанын және келесі практикада қай бағытқа көбірек көңіл бөлу қажет екенін анықтауға пайдаланылуы мүмкін.

Зерттеудің негізгі жаңалығы үш нәтижемен сипатталады:

1. Hard Skills және Soft Skills дағдыларын бағалауға арналған құрылымдық компоненттер мен критерийлер нақтыланды.

2. Педагогикалық практиканың 2 және 4 апталық кезеңдеріне арналған дағдылар картасы әзірленді.

3. Практикаға дейінгі және кейінгі диагностика арқылы дағдылардағы өзгеріс сипатталып, топтық рефлексия негізінде кәсіби дамудың басым бағыттары анықталды.

Практикалық тұрғыдан модельді педагогикалық колледждердің өндірістік және дипломалды практикасын ұйымдастыруда, тәлімгер мұғалімдер мен әдіскерлердің кері байланыс беруінде және болашақ мұғалімдердің кәсіби даму бағыттарын белгілеуде қолдануға болады.

Ұсыныстың басты ерекшелігі – педагогикалық практиканы тек өткізілген сабақтар жиынтығы ретінде емес, студенттің кәсіби және тұлғааралық дағдылары бір мезгілде көрінетін оқу-кәсіби орта ретінде қарастыру. Дағдылар картасы, өзін-өзі және өзара бағалау, педагогикалық бақылау мен топтық рефлексияны бірге қолдану әр студенттің кәсіби даму траекториясын нақты деректерге сүйеніп талқылауға мүмкіндік береді.

Әдебиеттер тізімі

1. Ağçam R. A study on the soft skills of pre-service teachers / R. Ağçam, A. Doğan // International Journal of Progressive Education. - 2021. - Vol. 17, № 4. - P. 35-48. <https://doi.org/10.29329/ijpe.2021.366.3>
2. Amorim C. Preservice physical education teachers' perceptions of initial teacher education / C. Amorim, E. Ribeiro-Silva // European Physical Education Review. - 2025. - Vol. 31, № 3. <https://doi.org/10.1177/1356336X241276825>
3. Macqual S. M. Assessing prospective teachers' soft skills curriculum implementation: Effects on teaching practicum success / S. M. Macqual, U. K. M. Salleh, H. Zulnaidi // South African Journal of Education. - 2021. - Vol. 41, № 3. - P. 1-21. <https://doi.org/10.15700/saje.v41n3a1915>
4. Mentari S. Student voices in teacher evaluation: The role of students in assessing student teachers' soft skills on Asistensi Mengajar Program / S. Mentari, U. Nuraini, R. Palil, V. Kitnuntaviwat // Journal of Accounting and Business Education. - 2026. - Vol. 10, № 4.
5. Heckman J. J. Hard evidence on soft skills / J. J. Heckman, T. Kautz // Labour Economics. - 2012. - Vol. 19, № 4. - P. 451-464. <https://doi.org/10.1016/j.labeco.2012.05.014>
6. Rusznyak L. What matters for competent teaching? A multinational comparison of teaching practicum assessment rubrics / L. Rusznyak, L. Österling // Teaching and Teacher Education. - 2024. - Vol. 151. - P. 104745. <https://doi.org/10.1016/j.tate.2024.104745>
7. Shkutina L. A. The need to form "hard skills" and "soft skills" for future pedagogues in the context of digitalization of education / L. A. Shkutina, T. G. Iskhakbayeva, Zh. A. Karmanova, J. Danek // Bulletin of the Karaganda University. Pedagogy Series. - 2022. - Vol. 108, № 4. - P. 7-13. <https://doi.org/10.31489/2022Ped4/7-13>
8. Казимова Д. А. Условия подготовки будущих учителей информатики на основе системного подхода / Д. А. Казимова // Bulletin of the Karaganda University. Pedagogy Series. - 2024. - Vol. 29, № 3. - P. 6-13. <https://doi.org/10.31489/2024ped3/6-13>

9. Есказинов, Ж. А. Формирование мягких навыков у будущих учителей иностранного языка / Ж. А. Есказинова, Л. С. Сырымбетова, А. К. Шаймерденова // Bulletin of the Karaganda University. Pedagogy Series. - 2025. - Vol. 30, № 1. - С. 30-37. <https://doi.org/10.31489/2025ped1/30-37>
10. OECD Learning Compass 2030. - OECD, 2019
11. Выготский Л. С. История развития высших психических функций / Л. С. Выготский // Собрание сочинений : в 6 т. - Москва : Педагогика, 1983. - Т. 3. - С. 5-328
12. Эльконин, Д. Б. К проблеме периодизации психического развития в детском возрасте / Д. Б. Эльконин // Вопросы психологии. - 1971. - № 4. - С. 6-20
13. Кузьмина Н. В. Профессионализм педагогической деятельности / Н. В. Кузьмина, А. А. Реан. - Санкт-Петербург ; Рыбинск : Научно-исследовательский центр развития творчества молодежи, 1993. - 54 с.
14. Жарықбаев Қ. Жантану негіздері : оқулық / Қ. Жарықбаев. - 4-басылым, өңделген. - Алматы : Кітап, 2002. - 415-416 б. - ISBN 9965-465-28-2
15. Panko E. A. The Problem of Psychological Development in Contemporary Developmental and Educational Psychology / E. A. Panko // Current Issues in Developmental and Educational Psychology. - Minsk : Vysheyschaya Shkola, 1980. - P. 3-24

У.З. Ешимова, Б.Б. Махмұтов, Ж.О. Оразымбетова, З.Н. Маханбетова

ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ И КРИТЕРИИ ОЦЕНИВАНИЯ ПРОФЕССИОНАЛЬНЫХ И МЯГКИХ НАВЫКОВ СТУДЕНТОВ ПЕДАГОГИЧЕСКОГО КОЛЛЕДЖА В ПРОЦЕССЕ ПЕДАГОГИЧЕСКОЙ ПРАКТИКИ

В статье рассматриваются педагогические условия и критерии оценки профессиональных навыков (hard skills) и гибких навыков (soft skills), формируемых у студентов педагогических колледжей в ходе педагогической практики. На основе проведенного исследования предложены критерии и показатели для оценки профессиональной компетентности студентов, а также их коммуникативных, организационных и рефлексивных навыков. Цель исследования - обосновать педагогические условия оценки данных навыков с учетом возрастных особенностей студентов, а также определить соответствующие критерии и показатели. В работе использовались такие методы, как анализ научной литературы, опросы, самооценка, взаимная оценка, педагогическое наблюдение и моделирование. Эмпирические данные собирались в 2024-2026 годах в педагогических колледжах Астаны и Кызылорды, в частности, в рамках специальности (01140100) - «Педагогика и методика начального обучения». В ходе исследования была разработана карта навыков для педагогической практики, а также определены структурные компоненты и критерии оценки как профессиональных (hard skills), так и надпрофессиональных (soft skills) навыков. Результаты самооценки, проведенной до и после педагогической практики, продемонстрировали положительные изменения в ряде профессиональных навыков студентов, а групповая рефлексия позволила определить конкретные направления для дальнейшего профессионального развития. Предложенные критерии обеспечивают возможность комплексной оценки профессионального и личностного потенциала студентов в период прохождения практики. Полученные данные служат основой для повышения уровня профессиональной готовности студентов и совершенствования процесса их адаптации к педагогической деятельности.

Ключевые слова: педагогическая практика, hard skills, soft skills, будущие учителя, оценивание, рефлексия, педагогический колледж.

U.Z. Yeshimova, B.B. Makhmutov, Zh.O. Orazymbetova, Z.N. Makhambetova

PEDAGOGICAL CONDITIONS AND CRITERIA FOR ASSESSING PROFESSIONAL HARD SKILLS AND SOFT SKILLS OF PEDAGOGICAL COLLEGE STUDENTS DURING TEACHING PRACTICE

The article examines the pedagogical conditions and assessment criteria for the professional qualities hard skills and soft skills developed by pedagogical college students during their teaching practicum. Based on the study conducted, the authors propose criteria and indicators for assessing students' professional competence, as well as their communicative, organizational, and reflective skills. The aim of the study is to substantiate the pedagogical conditions for assessing these skills, taking into account the age-related characteristics of students, and to define the corresponding criteria and indicators. The study employed methods such as the analysis of scientific literature, surveys, self-assessment, peer assessment, pedagogical observation, and modeling. Empirical data were collected between 2024 and 2026 at pedagogical colleges in Astana and Kyzylorda, specifically within the «Pedagogy and Methodology of Primary Education» (01140100) specialization. The study involved developing a skills matrix for teaching practicums and defining the structural components and assessment criteria for both professional (hard skills) and transferable (soft skills) competencies. The results of self-assessments conducted before and after the teaching practicum demonstrated positive changes in a number of the students' professional skills, while group reflection helped identify specific areas for further professional development. The proposed criteria enable a comprehensive assessment of the students' professional and personal potential during the practicum. The data obtained serve as a basis for enhancing the students' level of professional readiness and improving the process of their adaptation to teaching practice.

Keywords: pedagogical practice, hard skills, soft skills, future teachers, assessment, reflection, pedagogical college.

References

1. Ağçam, R., & Doğan, A. (2021). A study on the soft skills of pre-service teachers. *International Journal of Progressive Education*, 17(4), 35 - 48. <https://doi.org/10.29329/ijpe.2021.366.3>
2. Amorim, C., & Ribeiro-Silva, E. (2025). Preservice physical education teachers' perceptions of initial teacher education. *European Physical Education Review*, 31(3). <https://doi.org/10.1177/1356336X241276825>
3. Macqual, S. M., Salleh, U. K. M., & Zulnaidi, H. (2021). Assessing prospective teachers' soft skills curriculum implementation: Effects on teaching practicum success. *South African Journal of Education*, 41(3), 1 - 21. <https://doi.org/10.15700/saje.v41n3a1915>
4. Mentari, S., Nuraini, U., Palil, R., & Kitnuntaviwat, V. (2026). Student voices in teacher evaluation: The role of students in assessing student teachers' soft skills on Asistensi Mengajar Program. *Journal of Accounting and Business Education*, 10(4).
5. Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451 - 464. <https://doi.org/10.1016/j.labeco.2012.05.014>
6. Rusznyak, L., & Österling, L. (2024). What matters for competent teaching? A multinational comparison of teaching practicum assessment rubrics. *Teaching and Teacher Education*, 151, 104745. <https://doi.org/10.1016/j.tate.2024.104745>
7. Shkutina, L. A., Iskhakbayeva, T. G., Karmanova, Zh. A., & Danek, J. (2022). The need to form «hard skills» and «soft skills» for future pedagogues in the context of digitalization of education. *Bulletin of the Karaganda University. Pedagogy Series*, 108(4), 7 - 13. <https://doi.org/10.31489/2022Ped4/7-13>
8. Kazimova, D. A. (2024). *Usloviia podgotovki budushchikh uchitelei informatiki na osnove sistemnogo podkhoda* [Conditions for preparing future computer science teachers based on a systematic approach]. *Bulletin of the Karaganda University. Pedagogy Series*, 29(3), 6 - 13. <https://doi.org/10.31489/2024ped3/6-13>
9. Eskazinova, Zh. A., Syrymbetova, L. S., & Shaimerdenova, A. K. (2025). *Formirovanie miagkikh navykov u budushchikh uchitelei inostrannogo iazyka* [Formation of soft skills in future foreign language

teachers].. *Bulletin of the Karaganda University. Pedagogy Series*, 30(1), 30 - 37.
<https://doi.org/10.31489/2025ped1/30-37>

10. OECD. (2019). *OECD Learning Compass 2030*.

11. Vygotsky, L. S. (1983).). *Istoriia razvitiia vysshikh psikhicheskikh funktsii* [History of the development of higher mental functions]. In *Collected Works* (Vol. 3, pp. 5-328). Moscow: Pedagogika.

12. Elkonin, D. B. (1971).). *K probleme periodizatsii psikhicheskogo razvitiia v detskom vozraste* [On the problem of periodization of mental development in childhood]. *Questions of Psychology*, (4), 6 - 20.

13. Kuzmina, N. V., & Rean, A. A. (1993). *Professionalizm pedagogicheskoi deiatelnosti* [Professionalism of pedagogical activity]. St. Petersburg; Rybinsk: Research Center for the Development of Youth Creativity. 54 p.

14. Zharykbaev, K. (2002). *Zhantanu negizderi* [Fundamentals of psychology]. *Textbook* (4th revised ed.). Almaty: Kitap. pp. 415-416. ISBN 9965-465-28-2.

15. Panko E.A. «The Problem of Psychological Development in Contemporary Developmental and Educational Psychology» // *Current Issues in Developmental and Educational Psychology*. - Minsk: Vysheyskaya Shkola, 1980. - Pp. 3 - 24.

Авторлар туралы мәлімет:

Ешимова Ултай Зикировна (автор-корреспондент) - педагогика ғылымдарының кандидаты, Қорқыт ата атындағы Қызылорда Университетінің «Физика және математика» білім беру бағдарламасының аға оқытушысы, Қызылорда, Қазақстан Республикасы, ultai2005@gmail.com

Махмұтов Болат Бижанұлы - химия ғылымдарының кандидаты, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК Ғылыми зерттеу мен психометрика зертханасының бас ғылыми қызметкері, Астана қ., Қазақстан Республикасы, bolat200@mail.ru

Оразымбетова Жадыра Ормановна - педагогика ғылымдарының магистрі, М. Маметова атындағы Қызылорда педагогикалық жоғары колледжі, Қызылорда, Қазақстан Республикасы, zhadira2008@mail.ru

Маханбетова Зәуре Нағашыбайқызы - педагогика ғылымдарының магистрі, Астана халықаралық университеті жоғары педагогикалық колледжінің оқытушысы, Астана қ., Жансүгірова 4, zaure111@bk.ru

Сведения об авторах:

Ешимова Ултай Зикировна (автор-корреспондент) - кандидат педагогических наук, старший преподаватель образовательной программы «Физика и математика» Кызылординского университета имени Коркыт Ата, Кызылорда, Республика Казахстан, ultai2005@gmail.com

Махмұтов Болат Бижанұлы - кандидат химических наук, главный научный сотрудник лаборатории научных исследований и психометрики РГП на ПХВ «Национальный центр тестирования» МНВО РК, г. Астана, Республика Казахстан, bolat200@mail.ru

Оразымбетова Жадыра Ормановна - магистр педагогических наук, преподаватель атындағы Кызылординского высшего педагогического колледжа им. М. Маметовой, Кызылорда, Республика Казахстан, zhadira2008@mail.ru

Маханбетова Зәуре Нағашыбайқызы - магистр педагогических наук, преподаватель Высшего педагогического колледжа Международного университета Астана, г. Астана, ул. Жансүгірова 4, zaure111@bk.ru

Information about the authors:

Yeshimova Ultai Zikirovna (corresponding author) - Candidate of Pedagogical Sciences, Senior Lecturer of the «Physics and Mathematics» educational program at KorkytAta Kyzylorda University, Kyzylorda, Republic of Kazakhstan, ultai2005@gmail.com

Makhmutov Bolat Bizhanovich - Candidate of Chemical Sciences, Chief Researcher at the Laboratory of Scientific Research and Psychometrics «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Republic of Kazakhstan, bolat200@mail.ru

Orazymbetova Zhadira Ormanovna - Master of Pedagogical Sciences, Lecturer at the M. Mametova Kyzylorda Higher Pedagogical College, Kyzylorda, Republic of Kazakhstan, zhadira2008@mail.ru

Makhanbetova Zaure Nagashybaykyzy - Master of Pedagogical Sciences, Lecturer at the Higher Pedagogical College of the Astana International University, Astana, Zhansugirova Street 4, zaure111@bk.ru

Э.Ж. Шамшенова

Карагандинский национальный исследовательский университет имени академика Е.А. Букетова,
г. Караганды, Республика Казахстан
e-mail: shamshen7007@gmail.com

ORCID: 0009-0007-0457-8650

ЛИНГВОДИДАКТИЧЕСКИЕ ОСНОВЫ ФОРМИРОВАНИЯ ФРАЗОВОЙ РЕЧИ ПРИ ЗРИТЕЛЬНОЙ ДЕПРИВАЦИИ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ НАУЧНЫХ ПОДХОДОВ И МОДЕЛЬ КРИТЕРИАЛЬНОГО ОЦЕНИВАНИЯ

Статья посвящена сравнительному анализу научных подходов к формированию фразовой речи у детей со зрительной депривацией и обоснованию инструмента ее критериального оценивания. Актуальность исследования определяется противоречием: зарубежные работы фиксируют относительную сохранность структурных компонентов языка у слепых и слабовидящих детей, тогда как тифлопедагогическая традиция описывает системное недоразвитие фразы и вербализм. Цель - выявить причины этого расхождения и построить модель оценивания фразовой речи, освобожденную от зрительно зависимых источников ошибки измерения. Методология включает интегративный критический обзор публикаций из баз Scopus, Web of Science, PubMed и eLIBRARY с кодированием источников по шести критериям и сопоставлением пяти подходов: культурно-исторического, интеракционистского, нейрокогнитивного, функционально-прагматического и ассистивно-технологического; модель оценивания спроектирована в логике доказательно-центрированного дизайна (Evidence-Centered Design). Показано, что расхождение выводов объясняется различием изучаемых популяций, уровней анализа и стимульного материала: методики на сюжетных картинках смешивают языковой конструкт с состоянием зрения. Предложена трехслойная модель дефицита фразы (структурный, семантико-референтный, дискурсивно-прагматический слой) и карта критериального оценивания, включающая семь профилей речевых компетенций, полисенсорные стимулы в ситуациях социально-бытовой ориентировки, рубрику 0-3 балла, маркер перцептивной ошибки и интегральный индекс. Сформулирован план валидации инструмента (индекс содержательной валидности, внутриклассовая корреляция, модель Раша). Карта и карточки-задания предназначены логопедам и тифлопедагогам для входной, текущей и итоговой диагностики в специальных и инклюзивных организациях образования.

Ключевые слова: фразовая речь, зрительная депривация, слабовидящие младшие школьники, вербализм, критериальное оценивание, доказательно-центрированный дизайн, социально-бытовая ориентировка, лингводидактика.

Введение

Фразовая речь - уровень речевого развития, на котором ребенок впервые объединяет лексические, грамматические и смысловые средства в законченное высказывание. Для младшего школьника фраза является базовой единицей учебной коммуникации: из нее строятся ответ на вопрос учителя, пересказ, инструкция и описание. В культурно-исторической традиции высказывание рассматривается как движение от замысла через внутреннюю программу к развернутой речевой форме [1, 2]. Тифлопсихология уточняет, что нарушение зрения не меняет общих закономерностей речевого онтогенеза, но изменяет его чувственную опору: значение слова формируется на основе осязательного, слухового и двигательного опыта, а не зрительного образа [3].

Современная литература дает две несогласованные картины речевого развития при зрительной депривации. В исследовании В. Тадич и соавторов школьники с врожденными нарушениями зрения не отличались от зрячих сверстников по структурным языковым показателям, но уступали им в прагматике и социальной коммуникации [4]. Систематический обзор 2025 г. по 13 исследованиям пришел к выводу, что прямое

вмешательство на структурные компоненты языка обычно не требуется, а эффективными оказываются программы, направленные на взаимодействие в диаде «взрослый - ребенок» [5]. По данным родительских опросников, у слепых детей раннего возраста продуктивный словарь отстает примерно на полгода, однако его состав по частям речи и семантическим категориям совпадает с нормой [6]. В отечественной тифлопедагогике и логопедии, напротив, описываются бедность и малоинформативность фраз, неточное употребление лексики, опора на слово без представления об объекте; в констатирующем эксперименте автора (101 слабовидящий младший школьник) уровень сформированности фразовой речи оказался ниже среднего [7]. Популяционные данные также указывают, что у детей с нарушением зрения речевые и языковые нарушения диагностируются почти втрое чаще [8].

Для педагогических измерений это противоречие имеет принципиальный смысл. Если стимул диагностического задания требует зрительного анализа (сюжетная картинка, серия изображений), ошибка ребенка может отражать не языковую, а перцептивную трудность. В терминах теории валидности это конструктно-нерелевантная вариация: балл частично измеряет не то свойство, ради которого создан тест [9]. Систематический обзор практик массового оценивания учащихся с нарушениями зрения показывает, что доступность заданий для этих учащихся требует специальных технологических и методических решений [10]. Следовательно, расхождение выводов о «сохранности» и «недоразвитии» фразовой речи может быть во многом методологическим, и его разрешение требует соединения лингводидактики с теорией измерений.

Цель статьи - на основе сравнительного анализа научных подходов выявить причины расхождения данных о фразовой речи детей со зрительной депривацией и разработать модель и инструмент ее критериального оценивания для слабовидящих младших школьников. Исследование построено вокруг трех исследовательских вопросов (ИВ):

ИВ1. Какими теоретическими допущениями, единицами анализа и инструментами различаются основные подходы к фразовой речи при зрительной депривации?

ИВ2. Чем объясняется расхождение выводов о сохранности и о недоразвитии фразовой речи у детей с нарушениями зрения?

ИВ3. Какой должна быть модель оценивания фразовой речи слабовидящих младших школьников, минимизирующая влияние зрительного дефицита на результат измерения?

Научная новизна исследования определяется тем, что, в отличие от имеющихся систематических обзоров, ограниченных дошкольным возрастом [11] или англоязычными исследованиями из стран с высоким уровнем дохода [5], в едином аналитическом поле сопоставлены зарубежные парадигмы и отечественная тифлопедагогическая традиция, а выводы соотнесены с эмпирическими данными о слабовидящих младших школьниках Центральной Азии [7]. На этой основе: 1) обоснована трехслойная модель дефицита фразы (структурный, семантико-референтный и дискурсивно-прагматический слои), которая объясняет противоречие между данными о сохранности и о недоразвитии речи и переводит спор о «норме» в проверяемую гипотезу о том, какой слой фразы измеряет конкретная методика; 2) перцептивная ошибка введена как самостоятельная измерительная категория, а контрольная проба опознания - как процедура ее отделения от языковой ошибки, что операционализирует требование устранения конструктно-нерелевантной вариации [9] применительно к речевой диагностике слабовидящих детей; 3) разработана карта критериального оценивания, в которой семь профилей речевых компетенций соединены с тремя слоями фразы и с полисенсорными заданиями в ситуациях социально-бытовой ориентировки, а результатом измерения служит не единый балл, а профиль подындексов, позволяющий отличить вербализм от грамматического недоразвития.

Практическая значимость состоит в том, что карта оценивания, рубрика и банк карточек-заданий готовы к использованию логопедами и тифлопедагогами для входной, текущей и итоговой диагностики, профиль подындексов служит основанием для индивидуального маршрута коррекции, а план валидизации задает последовательность действий для превращения инструмента в стандартизованную методику.

Материалы и методы исследования

Дизайн. Использован интегративный критический обзор, который допускает совместный анализ эмпирических, обзорных и теоретических работ и потому подходит для области, где сосуществуют нейрокогнитивные эксперименты, клинические когортные исследования и педагогические концепции [12]. Обзор дополнен этапом проектирования оценочного инструмента по схеме доказательно-центрированного дизайна (ECD): модель конструктора - модель свидетельств - модель задания [13].

Источники и стратегия поиска. Поиск проводился в Scopus, Web of Science Core Collection, PubMed и eLIBRARY.RU. Англоязычная строка: («visual impairment» OR blind* OR «low vision») AND (language OR syntax OR narrative OR discourse OR pragmatic*) AND child*. Русскоязычные запросы сочетали термины «фразовая речь», «связная речь», «вербализм» с терминами «нарушение зрения», «слабовидящие», «тифлопедагогика». Основной период поиска - 2010-2026 гг.; классические работы, на которых основаны современные концепции, включались методом обратного прослеживания ссылок. Использованы также систематические обзоры 2015 и 2025 гг. как опорные карты области [5, 11].

Критерии отбора. Включались рецензируемые публикации, содержащие данные о лексике, синтаксисе, дискурсе или прагматике детей с нарушениями зрения либо о процедурах их оценки и коррекции; исследования взрослых с врожденной слепотой включались как модель компенсаторных механизмов. Исключались работы, посвященные только технике чтения по Брайлю, и публикации без описания методики. Отбор проходил в два этапа: по названиям и аннотациям, затем по полным текстам. Итоговый аналитический корпус составили 27 публикаций; 7 методологических работ по теории валидности, надежности и измерению использованы при проектировании инструмента.

Процедура анализа. Каждая публикация кодировалась по шести критериям (таблица 1). Дополнительно для каждого подхода экспертно оценивалась степень изученности семи профилей речевых компетенций, выделенных в диссертационном исследовании автора [7], по шкале 0-3 (0 - профиль не изучается; 3 - изучается системно, с описанной методикой). Результаты констатирующего эксперимента (101 слабовидящий младший школьник специализированных школ-интернатов городов Ташкента, Карши и Самарканда) использованы качественно - как эмпирическая проверка типологии ошибок, заложенной в модель. Рубрика оценивания нарративных умений строилась с учетом логики Narrative Scoring Scheme, психометрические свойства которой изучены на выборке младших школьников [14].

Таблица 1 - Критерии сравнительного анализа научных подходов

Код	Критерий	Что фиксировалось при кодировании
K1	Теоретическое основание	Исходная концепция речевого развития и роль зрения в ней
K2	Механизм нарушения	Первичный или вторичный характер дефицита; источник трудностей (опыт, взаимодействие, нейропластичность, доступ)
K3	Единица анализа	Слово, предложение, высказывание, текст, коммуникативный акт
K4	Стимулы и процедура оценки	Модальность стимула (зрительная, тактильная, слуховая, деятельностная); стандартизованность процедуры
K5	Направленность вмешательства	Ребенок, диада, среда, технические средства
K6	Уровень доказательности	Дизайн исследования, объем выборки, наличие контрольной группы
Примечание - Составлено автором; критерий K6 ориентирован на классификацию уровней доказательности, использованную в систематическом обзоре [5]		

Результаты и обсуждение

Сравнение научных подходов к фразовой речи при зрительной депривации

Отвечая на первый исследовательский вопрос, анализ корпуса позволил выделить пять подходов, которые различаются не столько выводами, сколько тем, что именно они считают единицей речи и где ищут причину трудностей (таблица 2).

Культурно-исторический (тифлопедагогический) подход рассматривает фразу как единицу речевого высказывания, порождаемого через внутреннее программирование [1, 2]. Трудности детей с нарушениями зрения трактуются как вторичные: сужение чувственного опыта разрывает связь слова и образа, что проявляется в вербализме [3]. Сильная сторона подхода - целостная модель порождения связного высказывания и развитая методика его диагностики; наиболее распространенные процедуры строятся на рассказе по сюжетной картинке и серии картинок [15]. Для слабовидящих детей именно это становится слабым местом: стимул требует того анализатора, который нарушен.

Интеракционистский подход переносит акцент с ребенка на диаду. У детей с нарушениями зрения затруднены совместное внимание и указательные жесты, что меняет структуру диалога и усвоение дейктических средств, в частности личных местоимений [16]. Лонгитюдные данные показывают, что характеристики взаимодействия матери с младенцем связаны с последующими когнитивными и речевыми исходами [17]; на этом основаны программы раннего домашнего вмешательства [18], тренинг родительской чуткости, проверенный в рандомизированном исследовании [19], и ранняя адаптация среды [20]. Фраза как самостоятельный объект оценки в этом подходе почти не выделяется.

Нейрокогнитивный (компенсаторный) подход показывает, что у людей с врожденной слепотой «зрительная» кора вовлекается в обработку предложений [21], взрослые слепые превосходят зрячих в понимании синтаксически сложных предложений [22], а такая перестройка имеет сенситивный период в детстве [23]. У слепых детей раннего возраста задержка словаря сочетается с нормативным составом лексикона [6]. Для лингводидактики отсюда следует: синтаксическая оформленность фразы не является неизбежным дефицитом зрительной депривации. Ограничение подхода - выборки взрослых с тотальной слепотой, что затрудняет перенос выводов на слабовидящих младших школьников.

Функционально-прагматический подход изучает употребление речи. Школьники с нарушениями зрения при сохранной грамматике испытывают трудности в социальной коммуникации [4]; в совместном рассказывании историй меняется использование лексики психических состояний [24]; затруднено понимание не прямых значений [25]; специфично употребление пространственного и временного дейксиса [26]. Прямое отношение к проблеме вербализма имеет исследование вербальных определений знакомых предметов: у слепых детей они отражают особый, преимущественно тактильно-функциональный перцептивный опыт [27]. Это позволяет переосмыслить вербализм - не как «пустоту» слова, а как иную сенсорную заземленность значения, которую нужно учитывать при оценке.

Ассистивно-технологический подход сосредоточен на доступе: альтернативная коммуникация и тактильные символы у детей с нарушением зрения и дополнительными нарушениями [28], доступность процедур массового оценивания [10], распространенность нарушений звукопроизношения [29] и раннего фонологического осознания [8]. Подход дает инструменты доступа, но не содержательную модель фразы.

Таблица 2 - Сравнительная матрица научных подходов к фразовой речи при зрительной депривации

Подход	Механизм трудностей (К1-К2)	Единица анализа (К3)	Оценка и стимулы (К4)	Вмешательство (К5)	Ограничение для слабовидящих младших школьников
Культурно-исторический [1-3, 15]	Вторичный дефицит: разрыв слова и образа, вербализм	Высказывание, связный текст	Рассказ по картинке и серии картинок	Ребенок: коррекционное обучение	Зрительно зависимые стимулы; нет норм для слабовидящих
Интеракционистский [16-20]	Дефицит совместного внимания и дейксиса в диаде	Коммуникативный акт	Наблюдение взаимодействия, опросники	Диада, семья, среда	Ранний возраст; фраза не выделена
Нейрокогнитивный [6, 21-23]	Компенсаторная пластичность; задержка словаря	Слово, предложение	Экспериментальные слуховые задачи, нейровизуализация	Не разрабатывается	Взрослые и totally слепые; перенос ограничен
Функционально-прагматический [4, 5, 24-27]	Прагматика, иная сенсорная основа значений	Дискурс, нарратив	Нарративные пробы, стандартизованные тесты	Диада, дискурсивная практика	Малые выборки, англоязычный материал
Ассистивно-технологический [8, 10, 28, 29]	Барьеры доступа к стимулу и ответу	Коммуникативное действие	Адаптированные форматы, ассистивные средства	Технические средства	Нет модели фразы как языковой единицы
Примечание - Составлено автором по результатам кодирования аналитического корпуса					

Сводная экспертная оценка изученности семи профилей речевых компетенций представлена на рисунке 1. Наиболее изучен профиль моно- и диалогической речи ($\Sigma = 9$), что отражает интерес зарубежных исследований к прагматике. Наименее изучены описательный ($\Sigma = 6$) и творческий ($\Sigma = 5$) рассказ; вне отечественной традиции они почти не рассматриваются. Ни один подход не охватывает все семь профилей с использованием незрительных стимулов; в этом состоит центральный пробел, на который направлена предлагаемая модель.

Подход	П1	П2	П3	П4	П5	П6	П7
Культурно-исторический	3	2	2	3	1	3	2
Интеракционистский	1	1	1	1	3	0	1
Нейрокогнитивный	1	3	1	0	0	0	0
Функционально-прагматический	2	1	2	3	3	2	2
Ассистивно-технологический	1	1	1	1	2	1	0
Σ по профилю	8	8	7	8	9	6	5

Рисунок 1 - Экспертная оценка изученности профилей речевых компетенций в рамках разных подходов (0-3 балла)

Примечание - П1 - смысловая целостность высказывания; П2 - синтаксическая оформленность; П3 - последовательность; П4 - связность; П5 - моно- и диалогическая речь; П6 - описательный рассказ; П7 - творческий рассказ. Шкала: 0 - не изучается; 1 - фрагментарно; 2 - частично; 3 - системно, с описанной методикой.

Причины расхождения выводов о фразовой речи

Второй исследовательский вопрос касается причин расхождения. Сопоставление подходов показывает, что противоречие между тезисами о сохранности и о недоразвитии фразовой речи имеет три источника.

Популяционный источник. Зарубежные выборки чаще включают детей с врожденной периферической слепотой без дополнительных нарушений, тогда как в специальных школах обучаются преимущественно слабовидящие дети с разнородной этиологией. Церебральные нарушения зрения сопряжены с когнитивными и нейроразвивающими последствиями, отличными от периферических [30]; уровень адаптивного поведения школьников различается в зависимости от степени нарушения зрения [31]; при сочетанных нарушениях траектория языкового развития меняется существенно [28]. Для воспитанников интернатов добавляется узость бытового опыта: дети отделены от приготовления пищи, ухода за одеждой и других повседневных действий, которые служат естественным источником речевых ситуаций [7].

Уровень анализа. Нейрокогнитивные и прагматические исследования измеряют в основном грамматику предложения и словарь, тогда как тифлопедагогическая традиция оценивает информативность и смысловую наполненность фразы. Эти данные не противоречат друг другу, а описывают разные слои одного явления. На этом основании предложена трехслойная модель дефицита фразы (рисунок 2): структурный слой (морфосинтаксис) относительно устойчив к зрительной депривации, семантико-референтный слой (соотнесение слова с объектом, признаком, действием и пространством) наиболее уязвим, дискурсивно-прагматический слой (последовательность, связность, учет слушателя) зависит от опыта взаимодействия.



Рисунок 2 - Трехслойная модель дефицита фразовой речи при зрительной депривации

Измерительный источник. Типичные ошибки, зафиксированные в констатирующем эксперименте, - ребенок не распознает действие персонажа на картинке, фраза распадается на два коротких предложения, ограничивается подлежащим и сказуемым без обстоятельств и определений [7] - не могут быть однозначно отнесены к речевым. Часть из них возникает на этапе опознания изображения, то есть до начала речевого программирования. Если такая

ошибка снижает речевой балл, возникает конструктивно-нерелевантная вариация [9], и итоговая оценка систематически занижает речевые возможности слабовидящего ребенка. С учетом данных о специфической сенсорной основе значений [27] вербализм также должен проверяться не только по словесному ответу, но и через пробу с реальным предметом.

Следовательно, расхождение выводов связано прежде всего с тем, что подходы измеряют разные слои фразы на разных популяциях и с разными по модальности стимулами, а не с различием самих детей. Наиболее управляемым из трех источников является измерительный - именно на него направлена модель, представленная ниже.

Модель и карта критериального оценивания фразовой речи

Ответом на третий исследовательский вопрос служит модель оценивания. Она построена по схеме ECD [13] и интегрирует блоки методики коррекционной работы, обоснованные в диссертационном исследовании автора (целевой, содержательный, организационно-процессуальный, результативно-оценочный) [7]. Ее исходные положения:

1) *конструкт* определяется как способность строить информативную, грамматически оформленную и контекстно уместную фразу в ситуации реальной деятельности;

2) *стимулы* - полисенсорные и деятельностные: реальный предмет, действие на занятии социально-бытовой ориентировки (СБО), звуковой сигнал; зрительный стимул допускается только в адаптированном виде и после контрольной пробы опознавания;

3) *оценка* ведется раздельно по трем слоям фразы, что позволяет получить не один балл, а диагностический профиль;

4) *перцептивная ошибка* фиксируется отдельным маркером и не снижает языковой балл.

Структура модели показана на рисунке 3.



Рисунок 3 - Модель критериального оценивания фразовой речи слабовидящих младших школьников, построенная по схеме ECD

Содержательное наполнение модели раскрывает карта критериального оценивания (таблица 3). Каждый профиль речевых компетенций проверяется заданием в ситуации СБО, а ответ оценивается по трем слоям фразы.

Таблица 3 - Карта критериального оценивания фразовой речи (фрагмент показателей)

Профиль	Задание и полисенсорный стимул (ситуация СБО)	Слой S: наблюдаемые показатели	Слой R: наблюдаемые показатели	Слой D: наблюдаемые показатели
П1. Смысловая целостность высказывания	Ответить одной фразой, что делает педагог (слышимое и осязаемое действие: наливает воду в стакан)	Полное предложение с подлежащим и сказуемым	Точное название действия и объекта	Фраза соответствует ситуации и вопросу
П2. Синтаксическая оформленность	Распространить фразу «Мальчик моет...» с опорой на предметы (мыло, полотенце)	Второстепенные члены; верное согласование и управление	Адекватные определения и обстоятельства	Порядок слов подчинен смыслу
П3. Последовательность	Выполнить и проговорить 3 шага: «помыть руки»	Сочинительная и подчинительная связь	Названы все значимые шаги	Шаги названы в логическом порядке
П4. Связность	Рассказать о только что выполненном действии (сервировка стола)	Средства межфразовой связи (потом, поэтому)	Нет подмены слов общими «это», «там»	Нет смысловых разрывов
П5. Моно- и диалогическая речь	Ролевая ситуация «В магазине» с реальными предметами	Вопросительные и побудительные конструкции	Уместная лексика ситуации	Инициатива, учет реплики собеседника
П6. Описательный рассказ	Описать предмет после тактильного обследования (зубная щетка, варежка)	Не менее 4 предложений разной структуры	Признаки из разных модальностей: форма, материал, функция	План описания: название - признаки - назначение
П7. Творческий рассказ	Придумать продолжение истории по звуковому началу (шум дождя, звонок в дверь)	Сложные предложения	Образная лексика, ментальные глаголы	Завязка - развитие - развязка
Примечание - Профили речевых компетенций приведены по [7]; распределение показателей по слоям S, R, D, полисенсорные стимулы и наблюдаемые показатели разработаны автором для настоящей модели				

Каждая из 21 оценочной точки (7 профилей × 3 слоя) оценивается по единой рубрике (таблица 4). Поведенческие дескрипторы уменьшают долю субъективного суждения эксперта и создают основу для проверки согласованности оценщиков.

Таблица 4 - Рубрика оценивания фразы по слоям

Таблица 1. Рубрика оценивания фразы по слоям			
Балл	Слой S (структурный)	Слой R (семантико-референтный)	Слой D (дискурсивно-прагматический)
3	Фраза полная, распространенная, без грамматических ошибок	Слова точно соотносены с объектом, признаком, действием; признаки из разных модальностей	Ответ последователен, связан, учитывает вопрос и собеседника
2	Фраза полная, 1-2 ошибки согласования или управления	Единичные неточные номинации, преобладают общие признаки	Незначительные пропуски звеньев, связь восстанавливается
1	Неполная или нераспространенная фраза (подлежащее + сказуемое)	Вербализм: слова без опоры на опыт, замены «это», «такое»	Фрагментарность, нарушена логика, требуется помощь
0	Отдельные слова или отказ	Номинация отсутствует или ошибочна	Ответ не соотносен с заданием
П	Маркер перцептивной ошибки: контрольная проба опознания не пройдена. Балл по слою R не выставляется, задание повторяется в доступной модальности (тактильной, слуховой, деятельностиной)		
Примечание - Составлено автором; логика дескрипторов слоя D согласована с принципами нарративной рубрики [14]			

Результаты сводятся в подындексы по слоям и интегральный индекс фразовой речи (ИФР):

$$I_S = \Sigma b_{kS} / 21 \cdot 100 \% ; \quad I_R = \Sigma b_{kR} / 21 \cdot 100 \% ; \quad I_D = \Sigma b_{kD} / 21 \cdot 100 \% ;$$

$$\text{ИФР} = (\Sigma_{k=1...7} \Sigma_{l=S, R, D} b_{kl}) / 63 \cdot 100 \% ,$$

где b_{kl} - балл (0-3) по профилю k и слою l . Предварительные пороги уровней: высокий - ИФР $\geq 75 \%$, средний - 50-74 %, недостаточный - 25-49 %, низкий - менее 25 %. Задания с маркером П после повторного предъявления оцениваются заново, поэтому знаменатель остается неизменным. Диагностическое значение имеет не только ИФР, но и соотношение подындексов: сочетание высокого I_S с низким I_R указывает на вербализм и требует работы с предметно-практическим опытом, а не с грамматикой.

Практическим носителем модели служит карточка-задание. Пример карточки для профиля П6 приведен в таблице 5; аналогичные карточки составляются для каждого профиля и каждой из четырех тем СБО, что дает банк из 28 базовых заданий.

Таблица 5 - Карточка-задание П6-Г1 «Описательный рассказ» (тема СБО «Личная гигиена»)

Код и профиль	П6-Г1. Описательный рассказ
Цель	Выявить умение построить описание предмета из 4 и более связанных фраз на основе полисенсорного обследования
Материал	Зубная щетка с рифленой ручкой, тюбик пасты, стакан; для слабовидящих - предметы контрастных цветов
Контрольная проба опознания	«Возьми предмет в руки. Что это?» Проба пройдена, если предмет назван или верно показано его назначение. Если нет - маркер П, повторное обследование с помощью педагога
Инструкция	«Расскажи о зубной щетке так, чтобы тот, кто ее никогда не держал, понял, какая она и для чего нужна»
Эталон ответа (3 балла по всем слоям)	«Это зубная щетка. У нее длинная пластмассовая ручка с бороздками, чтобы она не скользила. На конце жесткая щетинка. Щеткой чистят зубы утром и вечером. Сначала на щетинку выдавливают немного пасты»
Оценка S / R / D	S: число и разнообразие структуры фраз; R: признаки из разных модальностей (форма, материал, фактура, функция); D: план «название - признаки - назначение», средства связи
Типичные ошибки	Перечисление без фраз («ручка, щетка, зубы»); зрительные штампы без опыта («красивая, блестящая»); пропуск назначения
Коррекционный вывод	При $R \leq 1$ - упражнения «Учимся ощущать и называть»; при $D \leq 1$ - мнемосхема плана описания

Предложенный инструмент не может быть рекомендован к массовому использованию без эмпирической проверки. План валидации приведен в таблице 6: он охватывает содержательную валидность, надежность оценщиков, соответствие модели измерения и чувствительность к коррекционному воздействию.

Таблица 6 - План валидации карты критериального оценивания

Этап	Процедура	Критерий принятия	Обоснование
Содержательная валидность	Экспертиза заданий и дескрипторов (6-10 экспертов: логопеды, тифлопедагоги)	$I-CVI \geq 0,78$ для каждого задания	[32]
Надежность оценщиков	Независимая оценка аудиозаписей ответов двумя экспертами	$ICC \geq 0,75$ (хорошая согласованность)	[33]

Структурная валидность	Калибровка 21 оценочной точки по рейтинговой модели Раша	Infit/Outfit MnSq 0,6-1,4; надежность разделения испытуемых $\geq 0,80$	[34]
Конструктивная валидность	Сравнение баллов по заданиям с картинками и с полисенсорными стимулами	Различие баллов по слою R при сопоставимом слое S	[9]
Чувствительность	Сравнение экспериментальной и контрольной групп до и после коррекционной программы	Статистически значимый прирост ИФР в экспериментальной группе	[7]

Обсуждение. Полученные результаты согласуются с выводом систематического обзора о том, что прямая тренировка грамматики для детей с нарушениями зрения обычно не является приоритетом [5]. Однако для слабовидящих младших школьников, обучающихся в условиях интерната, акцент смещается: мишенью становятся семантико-референтный и дискурсивный слои фразы, а естественной средой их формирования - занятия СБО, где слово возникает внутри практического действия. Тем самым идея диадического подхода о значимости совместной деятельности [17, 18] переносится из семьи раннего возраста в школьный коррекционный процесс. При этом модель не противопоставляет себя отечественной методике рассказа по картинке [15]: картинка сохраняется как адаптированный стимул, но после контрольной пробы опознания, что делает интерпретацию балла корректной.

Исследование имеет ограничения. Кодирование источников и экспертная оценка изученности профилей выполнены одним автором; для последующей публикации запланирована независимая перекодировка вторым экспертом. Выводы нейрокогнитивных работ получены на взрослых с тотальной слепотой и переносятся на слабовидящих детей с осторожностью. Пороги уровней ИФР носят предварительный характер и подлежат уточнению после калибровки по модели Раша.

Заключение

Сравнительный анализ показал, что пять подходов к фразовой речи при зрительной депривации различаются прежде всего единицей анализа и тем, где они ищут причину трудностей. Культурно-исторический подход изучает высказывание и объясняет трудности разрывом слова и образа, интеракционистский - коммуникативный акт в диаде, нейрокогнитивный - слово и предложение, фиксируя компенсаторную сохранность синтаксиса, функционально-прагматический - дискурс и иную сенсорную основу значений, ассистивно-технологический - барьеры доступа. Ни один из подходов не охватывает все семь профилей речевых компетенций с опорой на незрительные стимулы.

Противоречие между данными о сохранности и о недоразвитии фразовой речи снимается, если учитывать различие изучаемых популяций, уровней анализа и модальности стимулов. В рамках трехслойной модели эти данные согласуются: структурный слой фразы относительно устойчив к зрительной депривации, семантико-референтный наиболее уязвим, дискурсивно-прагматический зависит от опыта взаимодействия. Для слабовидящих младших школьников, обучающихся в интернатах, основной мишенью коррекции становятся именно второй и третий слои.

Оценивание фразовой речи слабовидящих детей корректно только при условии, что языковой конструкт отделен от перцептивного доступа к стимулу. Предложенная карта критериального оценивания обеспечивает это за счет полисенсорных заданий в ситуациях социально-бытовой ориентировки, контрольной пробы опознания, отдельной оценки трех слоев фразы и расчета подындексов I_S , I_R , I_D и интегрального индекса ИФР. Дальнейшая работа связана с реализацией плана валидизации - экспертизой содержательной валидности, проверкой согласованности оценщиков и калибровкой по модели Раша, - а также с адаптацией карты для казахско- и узбекоязычных учащихся.

Список литературы

1. Выготский Л.С. Мышление и речь. - 5-е изд., испр. - М.: Лабиринт, 1999. - 352 с.
2. Лурия А.Р. Язык и сознание. - М.: Изд-во Московского университета, 1979. - 320 с.
3. Литвак А.Г. Психология слепых и слабовидящих. - СПб.: Изд-во РГПУ им. А.И. Герцена, 1998. - 271 с.
4. Tadić V., Pring L., Dale N. Are language and social communication intact in children with congenital visual impairment at school age? // *Journal of Child Psychology and Psychiatry*. - 2010. - Vol. 51, No. 6. - P. 696-705. - <https://doi.org/10.1111/j.1469-7610.2009.02200.x>
5. Micheletti S., Dusi L., Zordan T. et al. Understanding the effect of visual impairment on children's language skills and strategies for promotion: A systematic review // *Journal of Visual Impairment & Blindness*. - 2025. - Vol. 119, No. 5. - P. 344-363. - <https://doi.org/10.1177/0145482X251377471>
6. Campbell E., Casillas R., Bergelson E. The role of vision in the acquisition of words: Vocabulary development in blind toddlers // *Developmental Science*. - 2024. - Art. e13475. - <https://doi.org/10.1111/desc.13475>
7. Шамшенова Э.Ж. Усовершенствование методики коррекционной работы по развитию фразовой речи: автореф. дис. ... д-ра философии (PhD) по пед. наукам: 13.00.03 - Специальная педагогика. - Ташкент: Национальный педагогический университет Узбекистана, 2026.
8. Brouwer K.K., Gordon-Pershey M., Stransky M. Speech, language, and literacy in children with visual impairments: The National Survey of Children's Health // *Communication Disorders Quarterly*. - 2024. - Vol. 45, No. 4. - P. 263-269. - <https://doi.org/10.1177/15257401231184419>
9. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning // *American Psychologist*. - 1995. - Vol. 50, No. 9. - P. 741-749. - <https://doi.org/10.1037/0003-066X.50.9.741>
10. Leria L.A., Benitez P., Fraga F.J. Assistive technology in large-scale assessments for students with visual impairments: A systematic review and recommendations based on the Brazilian reality // *Education and Information Technologies*. - 2021. - Vol. 26, No. 3. - P. 3543-3573. - <https://doi.org/10.1007/s10639-020-10419-6>
11. Mosca R., Kritzinger A., van der Linde J. Language and communication development in preschool children with visual impairment: A systematic review // *South African Journal of Communication Disorders*. - 2015. - Vol. 62, No. 1. - P. e1-e10. - <https://doi.org/10.4102/sajcd.v62i1.119>
12. Whittemore R., Knafl K. The integrative review: Updated methodology // *Journal of Advanced Nursing*. - 2005. - Vol. 52, No. 5. - P. 546-553. - <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
13. Mislevy R.J., Steinberg L.S., Almond R.G. On the structure of educational assessments // *Measurement: Interdisciplinary Research and Perspectives*. - 2003. - Vol. 1, No. 1. - P. 3-62. - https://doi.org/10.1207/S15366359MEA0101_02
14. Heilmann J., Miller J.F., Nockerts A., Dunaway C. Properties of the narrative scoring scheme using narrative retells in young school-age children // *American Journal of Speech-Language Pathology*. - 2010. - Vol. 19, No. 2. - P. 154-166. - [https://doi.org/10.1044/1058-0360\(2009/08-0024\)](https://doi.org/10.1044/1058-0360(2009/08-0024))
15. Глухов В.П. Формирование связной речи детей дошкольного возраста с общим речевым недоразвитием. - 2-е изд., испр. и доп. - М.: АРКТИ, 2004. - 168 с.
16. Pérez-Pereira M. Deixis, personal reference, and the use of pronouns by blind children // *Journal of Child Language*. - 1999. - Vol. 26, No. 3. - P. 655-680. - <https://doi.org/10.1017/S0305000999003955>
17. Sakkalou E., O'Reilly M.A., Sakki H. et al. Mother-infant interactions with infants with congenital visual impairment and associations with longitudinal outcomes in cognition and language // *Journal of Child Psychology and Psychiatry*. - 2021. - Vol. 62, No. 6. - P. 742-750. - <https://doi.org/10.1111/jcpp.13308>
18. Dale N.J., Sakkalou E., O'Reilly M.A. et al. Home-based early intervention in infants and young children with visual impairment using the Developmental Journal: Longitudinal cohort study // *Developmental Medicine & Child Neurology*. - 2019. - Vol. 61, No. 6. - P. 697-709. - <https://doi.org/10.1111/dmcn.14081>
19. Platje E., Sterkenburg P., Overbeek M. et al. The efficacy of VIPP-V parenting training for parents of young children with a visual or visual-and-intellectual disability: A randomized controlled trial // *Attachment & Human Development*. - 2018. - Vol. 20, No. 5. - P. 455-472. - <https://doi.org/10.1080/14616734.2018.1428997>

20. Fazzi E., Micheletti S., Calza S. et al. Early visual training and environmental adaptation for infants with visual impairment // *Developmental Medicine & Child Neurology*. - 2021. - Vol. 63, No. 10. - P. 1180-1193. - <https://doi.org/10.1111/dmcn.14865>
21. Bedny M., Pascual-Leone A., Dodell-Feder D. et al. Language processing in the occipital cortex of congenitally blind adults // *Proceedings of the National Academy of Sciences of the USA*. - 2011. - Vol. 108, No. 11. - P. 4429-4434. - <https://doi.org/10.1073/pnas.1014818108>
22. Loiotile R., Lane C., Omaki A., Bedny M. Enhanced performance on a sentence comprehension task in congenitally blind adults // *Language, Cognition and Neuroscience*. - 2020. - Vol. 35, No. 8. - P. 1010-1023. - <https://doi.org/10.1080/23273798.2019.1706753>
23. Pant R., Kanjlia S., Bedny M. A sensitive period in the neural phenotype of language in blind individuals // *Developmental Cognitive Neuroscience*. - 2020. - Vol. 41. - Art. 100744. - <https://doi.org/10.1016/j.dcn.2019.100744>
24. Tadić V., Pring L., Dale N. Story discourse and use of mental state language between mothers and school-aged children with and without visual impairment // *International Journal of Language & Communication Disorders*. - 2013. - Vol. 48, No. 6. - P. 679-688. - <https://doi.org/10.1111/1460-6984.12040>
25. Pijnacker J., Vervloed M.P.J., Steenbergen B. Pragmatic abilities in children with congenital visual impairment: An exploration of non-literal language and advanced theory of mind understanding // *Journal of Autism and Developmental Disorders*. - 2012. - Vol. 42, No. 11. - P. 2440-2449. - <https://doi.org/10.1007/s10803-012-1500-5>
26. Iossifova R., Marmolejo-Ramos F. When the body is time: Spatial and temporal deixis in children with visual impairments and sighted children // *Research in Developmental Disabilities*. - 2013. - Vol. 34, No. 7. - P. 2173-2184. - <https://doi.org/10.1016/j.ridd.2013.03.030>
27. Vinter A., Fernandes V., Orlandi O., Morgan P. Verbal definitions of familiar objects in blind children reflect their peculiar perceptual experience // *Child: Care, Health and Development*. - 2013. - Vol. 39, No. 6. - P. 856-863. - <https://doi.org/10.1111/cch.12002>
28. Philip V.S., Suryasree P.K., Santhosh K.E. Efficacy and social validity of intervention using speech generating device and tactile symbols to develop communication and language in a child with visual impairment and additional disability // *Disability and Rehabilitation: Assistive Technology*. - 2025. - Vol. 20, No. 7. - P. 2557-2572. - <https://doi.org/10.1080/17483107.2025.2539436>
29. Gordon-Pershey M., Zeszut S., Brouwer K. A survey of speech sound productions in children with visual impairments // *Communication Disorders Quarterly*. - 2019. - Vol. 40, No. 4. - P. 206-219. - <https://doi.org/10.1177/1525740118789101>
30. Chokron S., Dutton G.N. From vision to cognition: Potential contributions of cerebral visual impairment to neurodevelopmental disorders // *Journal of Neural Transmission*. - 2023. - Vol. 130, No. 3. - P. 409-424. - <https://doi.org/10.1007/s00702-022-02572-8>
31. Bathelt J., de Haan M., Dale N.J. Adaptive behaviour and quality of life in school-age children with congenital visual disorders and different levels of visual impairment // *Research in Developmental Disabilities*. - 2019. - Vol. 85. - P. 154-162. - <https://doi.org/10.1016/j.ridd.2018.12.003>
32. Lynn M.R. Determination and quantification of content validity // *Nursing Research*. - 1986. - Vol. 35, No. 6. - P. 382-385. - <https://doi.org/10.1097/00006199-198611000-00017>
33. Koo T.K., Li M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research // *Journal of Chiropractic Medicine*. - 2016. - Vol. 15, No. 2. - P. 155-163. - <https://doi.org/10.1016/j.jcm.2016.02.012>
34. Bond T.G., Yan Z., Heene M. Applying the Rasch model: Fundamental measurement in the human sciences. - 4th ed. - New York: Routledge, 2020. - 376 p. - <https://doi.org/10.4324/9780429030499>

Э.Ж. Шамшенова**КӨРУ ДЕПРИВАЦИЯСЫ ЖАҒДАЙЫНДА ФРАЗАЛЫҚ СӨЙЛЕУДІ
ҚАЛЫПТАСТЫРУДЫҢ ЛИНГВОДИДАКТИКАЛЫҚ НЕГІЗДЕРІ: ҒЫЛЫМИ
ТӘСІЛДЕРДІ САЛЫСТЫРМАЛЫ ТАЛДАУ ЖӘНЕ КРИТЕРИАЛДЫ БАҒАЛАУ
МОДЕЛІ**

Мақала көру депривациясы бар балалардың фразалық сөйлеуін қалыптастыруға қатысты ғылыми тәсілдерді салыстырмалы талдауға және оны критериялды бағалау құралын негіздеуге арналған. Зерттеудің өзектілігі мынадай қайшылықпен анықталады: шетелдік зерттеулер соқыр және нашар көретін балаларда тілдің құрылымдық компоненттерінің салыстырмалы түрде сақталатынын көрсетсе, тифлопедагогикалық дәстүр фразаның жүйелі дамымауын және вербализмді сипаттайды. Мақсаты - осы алшақтықтың себептерін анықтау және өлшеу қатесінің көруге тәуелді көздерінен арылтылған фразалық сөйлеуді бағалау моделін құру. Әдіснамасы Scopus, Web of Science, PubMed және eLIBRARY деректер қорларындағы жарияланымдарға интегративті сыни шолу жасауды, дереккөздерді алты критерий бойынша кодтауды және бес тәсілді - мәдени-тарихи, интеракционистік, нейрокогнитивтік, функционалдық-прагматикалық және ассистивтік-технологиялық тәсілдерді салыстыруды қамтиды; бағалау моделі дәлелдерге негізделген жобалау (Evidence-Centered Design) қағидаттары бойынша әзірленді. Қорытындылардағы алшақтық зерттелген популяциялардың, талдау деңгейлерінің және стимулдық материалдың әртүрлілігімен түсіндірілетіні көрсетілді: сюжетті суреттерге негізделген әдістемелер тілдік конструктіні көру жағдайымен араластырып жібереді. Фраза тапшылығының үш қабатты моделі (құрылымдық, семантикалық-референттік, дискурсивтік-прагматикалық қабаттар) және сөйлеу құзыреттерінің жеті профилін, әлеуметтік-тұрмыстық бағдарлау жағдаяттарындағы полисенсорлық стимулдарды, 0-3 балдық рубриканы, перцептивтік қате белгісін және интегралдық индексті қамтитын критериялды бағалау картасы ұсынылды. Құралды валидизациялау жоспары (мазмұндық валидтілік индексі, класс ішілік корреляция, Раш моделі) тұжырымдалды. Карта мен тапсырма-карточкалар логопедтер мен тифлопедагогтарға арнайы және инклюзивті білім беру ұйымдарында кіріс, ағымдағы және қорытынды диагностика жүргізуге арналған.

Түйін сөздер: фразалық сөйлеу, көру депривациясы, нашар көретін бастауыш сынып оқушылары, вербализм, критериялды бағалау, дәлелдерге негізделген жобалау, әлеуметтік-тұрмыстық бағдарлау, лингводидактика.

E.Zh. Shamshenova**LINGUODIDACTIC FOUNDATIONS OF PHRASAL SPEECH DEVELOPMENT UNDER
VISUAL DEPRIVATION: A COMPARATIVE ANALYSIS OF SCIENTIFIC
APPROACHES AND A CRITERION-REFERENCED ASSESSMENT MODEL**

The article presents a comparative analysis of scientific approaches to the development of phrasal speech in children with visual deprivation and substantiates an instrument for its criterion-referenced assessment. It addresses a contradiction: international research reports relative preservation of structural language components in blind children and children with low vision, whereas the typhlopedagogical tradition describes systemic underdevelopment of the phrase and verbalism. The aim is to identify the causes of this discrepancy and to design an assessment model for phrasal speech that is free from vision-dependent sources of measurement error. The methodology combines an integrative critical review of publications from Scopus, Web of Science, PubMed and eLIBRARY, coding of sources against six criteria and a comparison of five approaches: cultural-historical, interactionist, neurocognitive, functional-pragmatic and assistive-technological. The assessment model follows Evidence-Centered Design. The divergence is shown to stem from differences in populations, levels of analysis and stimulus materials: picture-based procedures confound the linguistic construct with the child's visual status. A three-layer model of phrase deficit (structural, semantic-referential and discourse-pragmatic layers) is proposed, together with an assessment map that includes seven speech competence profiles, multisensory stimuli from social and everyday living orientation lessons, a 0-3 scoring rubric, a perceptual error flag and an integral index. A validation plan (content validity

index, intraclass correlation, Rasch modelling) is outlined. The map and task cards support baseline, formative and summative assessment by speech therapists and teachers of the visually impaired in special and inclusive schools.

Keywords: phrasal speech, visual deprivation, primary school children with low vision, verbalism, criterion-referenced assessment, evidence-centered design, social and everyday living orientation, linguodidactics.

References

1. Vygotskii, L. S. (1999). *Myshlenie i rech* [Thinking and speech] (5th ed.). Labirint.
2. Lurii, A. R. (1979). *Iazyk i soznanie* [Language and consciousness]. Izdatelstvo Moskovskogo universiteta.
3. Litvak, A. G. (1998). *Psikhologiya slepykh i slabovidiashchikh* [Psychology of the blind and visually impaired]. Izdatelstvo RGPU im. A. I. Gertsena.
4. Tadić, V., Pring, L., & Dale, N. (2010). Are language and social communication intact in children with congenital visual impairment at school age? *Journal of Child Psychology and Psychiatry*, 51(6), 696-705. <https://doi.org/10.1111/j.1469-7610.2009.02200.x>
5. Micheletti, S., Dusi, L., Zordan, T., Merabet, L. B., & Fazzi, E. (2025). Understanding the effect of visual impairment on children's language skills and strategies for promotion: A systematic review. *Journal of Visual Impairment & Blindness*, 119(5), 344-363. <https://doi.org/10.1177/0145482X251377471>
6. Campbell, E., Casillas, R., & Bergelson, E. (2024). The role of vision in the acquisition of words: Vocabulary development in blind toddlers. *Developmental Science*, Article e13475. <https://doi.org/10.1111/desc.13475>
7. Shamsheva, E. Zh. (2026). *Uovershenstvovanie metodiki korrektsionnoi raboty po razvitiyu frazovoi rechi* [Improving the methodology of corrective work on the development of phrasal speech] [Extended abstract of PhD dissertation]. Natsionalnyi pedagogicheskii universitet Uzbekistana.
8. Brouwer, K. K., Gordon-Pershey, M., & Stransky, M. (2024). Speech, language, and literacy in children with visual impairments: The National Survey of Children's Health. *Communication Disorders Quarterly*, 45(4), 263-269. <https://doi.org/10.1177/15257401231184419>
9. Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
10. Leria, L. A., Benitez, P., & Fraga, F. J. (2021). Assistive technology in large-scale assessments for students with visual impairments: A systematic review and recommendations based on the Brazilian reality. *Education and Information Technologies*, 26(3), 3543-3573. <https://doi.org/10.1007/s10639-020-10419-6>
11. Mosca, R., Kritzing, A., & van der Linde, J. (2015). Language and communication development in preschool children with visual impairment: A systematic review. *South African Journal of Communication Disorders*, 62(1), e1-e10. <https://doi.org/10.4102/sajcd.v62i1.119>
12. Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, 52(5), 546-553. <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
13. Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62. https://doi.org/10.1207/S15366359MEA0101_02
14. Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology*, 19(2), 154-166. [https://doi.org/10.1044/1058-0360\(2009/08-0024\)](https://doi.org/10.1044/1058-0360(2009/08-0024))
15. Glukhov, V. P. (2004). *Formirovanie svyaznoi rechi detei doshkolnogo vozrasta s obshchim rechevym nedorazvitiem* [Formation of coherent speech in preschool children with general speech underdevelopment] (2nd ed.). ARKTI.
16. Pérez-Pereira, M. (1999). Deixis, personal reference, and the use of pronouns by blind children. *Journal of Child Language*, 26(3), 655-680. <https://doi.org/10.1017/S0305000999003955>
17. Sakalou, E., O'Reilly, M. A., Sakki, H., Springall, C., de Haan, M., Salt, A. T., & Dale, N. J. (2021). Mother-infant interactions with infants with congenital visual impairment and associations with longitudinal outcomes in cognition and language. *Journal of Child Psychology and Psychiatry*, 62(6), 742-750. <https://doi.org/10.1111/jcpp.13308>

18. Dale, N. J., Sakkalou, E., O'Reilly, M. A., Springall, C., Sakki, H., Glew, S., Pissaridou, E., De Haan, M., & Salt, A. T. (2019). Home-based early intervention in infants and young children with visual impairment using the Developmental Journal: Longitudinal cohort study. *Developmental Medicine & Child Neurology*, 61(6), 697-709. <https://doi.org/10.1111/dmcn.14081>
19. Platje, E., Sterkenburg, P., Overbeek, M., Kef, S., & Schuengel, C. (2018). The efficacy of VIPP-V parenting training for parents of young children with a visual or visual-and-intellectual disability: A randomized controlled trial. *Attachment & Human Development*, 20(5), 455-472. <https://doi.org/10.1080/14616734.2018.1428997>
20. Fazzi, E., Micheletti, S., Calza, S., Merabet, L., Rossi, A., Galli, J., & Early Visual Intervention Study Group. (2021). Early visual training and environmental adaptation for infants with visual impairment. *Developmental Medicine & Child Neurology*, 63(10), 1180-1193. <https://doi.org/10.1111/dmcn.14865>
21. Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences*, 108(11), 4429-4434. <https://doi.org/10.1073/pnas.1014818108>
22. Loiotile, R., Lane, C., Omaki, A., & Bedny, M. (2020). Enhanced performance on a sentence comprehension task in congenitally blind adults. *Language, Cognition and Neuroscience*, 35(8), 1010-1023. <https://doi.org/10.1080/23273798.2019.1706753>
23. Pant, R., Kanjlia, S., & Bedny, M. (2020). A sensitive period in the neural phenotype of language in blind individuals. *Developmental Cognitive Neuroscience*, 41, Article 100744. <https://doi.org/10.1016/j.dcn.2019.100744>
24. Tadić, V., Pring, L., & Dale, N. (2013). Story discourse and use of mental state language between mothers and school-aged children with and without visual impairment. *International Journal of Language & Communication Disorders*, 48(6), 679-688. <https://doi.org/10.1111/1460-6984.12040>
25. Pijnacker, J., Vervloed, M. P. J., & Steenbergen, B. (2012). Pragmatic abilities in children with congenital visual impairment: An exploration of non-literal language and advanced theory of mind understanding. *Journal of Autism and Developmental Disorders*, 42(11), 2440-2449. <https://doi.org/10.1007/s10803-012-1500-5>
26. Iossifova, R., & Marmolejo-Ramos, F. (2013). When the body is time: Spatial and temporal deixis in children with visual impairments and sighted children. *Research in Developmental Disabilities*, 34(7), 2173-2184. <https://doi.org/10.1016/j.ridd.2013.03.030>
27. Vinter, A., Fernandes, V., Orlandi, O., & Morgan, P. (2013). Verbal definitions of familiar objects in blind children reflect their peculiar perceptual experience. *Child: Care, Health and Development*, 39(6), 856-863. <https://doi.org/10.1111/cch.12002>
28. Philip, V. S., Suryasree, P. K., & Santhosh, K. E. (2025). Efficacy and social validity of intervention using speech generating device and tactile symbols to develop communication and language in a child with visual impairment and additional disability. *Disability and Rehabilitation: Assistive Technology*, 20(7), 2557-2572. <https://doi.org/10.1080/17483107.2025.2539436>
29. Gordon-Pershey, M., Zeszut, S., & Brouwer, K. (2019). A survey of speech sound productions in children with visual impairments. *Communication Disorders Quarterly*, 40(4), 206-219. <https://doi.org/10.1177/1525740118789101>
30. Chokron, S., & Dutton, G. N. (2023). From vision to cognition: Potential contributions of cerebral visual impairment to neurodevelopmental disorders. *Journal of Neural Transmission*, 130(3), 409-424. <https://doi.org/10.1007/s00702-022-02572-8>
31. Bathelt, J., de Haan, M., & Dale, N. J. (2019). Adaptive behaviour and quality of life in school-age children with congenital visual disorders and different levels of visual impairment. *Research in Developmental Disabilities*, 85, 154-162. <https://doi.org/10.1016/j.ridd.2018.12.003>
32. Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-385. <https://doi.org/10.1097/00006199-198611000-00017>
33. Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
34. Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. <https://doi.org/10.4324/9780429030499>

Автор туралы мәлімет:

Шамшенова Эльмира Жаксылыковна - педагогика ғылымдарының магистрі, академик Е.А. Бөкетов атындағы Қарағанды ұлттық зерттеу университетінің арнайы және инклюзивті білім беру кафедрасының аға оқытушысы, Қарағанды, Қазақстан Республикасы, e-mail: shamshen7007@gmail.com

Сведения об авторе:

Шамшенова Эльмира Жаксылыковна - магистр педагогических наук, старший преподаватель кафедры специального и инклюзивного образования Карагандинского национального исследовательского университета имени академика Е.А. Букетова, Караганды, Республика Казахстан, e-mail: shamshen7007@gmail.com

Information about author:

Shamshenova Elmira Zhaksilykovna - Master of Pedagogical Sciences, Senior Lecturer at the Department of Special and Inclusive Education, Karaganda Buketov National Research University, Karagandy, Republic of Kazakhstan, e-mail: shamshen7007@gmail.com

Ж.Ж. Бейсенова*¹, Л.Е. Шинетова², Г.К. Жабаетова³, С.Ж. Анафияева⁴^{1,2,3,4}РГП на ПХВ «Национальный центр тестирования» МНВО РК, г. Астана, Республика Казахстан

*e-mail: zhanar-beisenova@mail.ru

¹ORCID 0000-0002-7347-1538, ²ORCID 0000-0003-4280-7999,³ORCID 0009-0006-1003-6316, ⁴ORCID 0009-0000-9124-9976

ДИФФЕРЕНЦИАЛЬНОЕ ФУНКЦИОНИРОВАНИЕ ТЕСТОВЫХ ЗАДАНИЙ В ТЕСТИРОВАНИИ ПЕДАГОГОВ НА ОСНОВЕ МОДЕЛИ РАША

В статье представлены результаты психометрического анализа дифференциального функционирования заданий (DIF) по признаку пола на материале трех параллельных вариантов теста для педагогов. Цель исследования - оценить наличие, величину и направление DIF-эффектов в рамках модели Раша и определить их статистическую устойчивость при контроле множественных сравнений. Предполагалось, что при сопоставимой трудности параллельных вариантов доля и направление DIF-эффектов будут ограничены отдельными заданиями и не образуют системного смещения в пользу одного пола. Материалом послужили результаты тестирования 1417 испытуемых (1071 женщина, 346 мужчин), выполнявших три варианта теста по 50 дихотомически оцениваемых заданий каждый; обработка данных проведена в программе Winsteps 5.8.3.0. Дифференциальное функционирование оценивалось по величине DIF, с дополнительной проверкой по статистике Mantel-Haenszel и контролем ложных обнаружений (FDR) по процедуре Бенджамини-Хохберга, примененной отдельно в каждом варианте. По первичному критерию DIF выявлен в 25 из 150 позиций «задание × вариант» (16,7%); после FDR-коррекции статистически устойчивыми остались 9 позиций (6,0%). Направление устойчивых эффектов оказалось практически сбалансированным: 4 задания труднее для женщин, 5 - для мужчин, что не дает оснований говорить о системном смещении трудности в пользу одного пола. Наибольшая концентрация DIF-сигналов отмечена в варианте 703. Результаты в целом подтверждают приемлемый уровень измерительной инвариантности использованных вариантов теста по признаку пола и позволяют сформулировать рекомендации по регулярному мониторингу DIF при разработке и эксплуатации параллельных форм теста.

Ключевые слова: дифференциальное функционирование заданий (DIF), модель Раша, инвариантность измерения, тестирование педагогов, гендерные различия, статистика Mantel-Haenszel.

Введение

Справедливость и сопоставимость результатов стандартизированного тестирования педагогов напрямую зависят от того, насколько тестовые задания одинаково функционируют для различных подгрупп испытуемых при одном и том же уровне измеряемой способности. Это свойство в литературе по психометрическому анализу принято называть инвариантностью измерения, а его нарушение на уровне отдельного задания - дифференциальным функционированием задания, или DIF (Differential Item Functioning) [1, 2]. Модель Раша, лежащая в основе большинства современных систем педагогических измерений, включая аналитическую платформу Winsteps, задает естественную основу для проверки инвариантности: если параметр трудности задания статистически различается для двух групп испытуемых, сопоставимых по общему уровню способности, это рассматривается как индикатор возможного DIF [1, 3].

Начиная с работ Holland и Thayer, статистика Mantel-Haenszel стала одним из наиболее распространенных непараметрических инструментов обнаружения DIF наряду с методами, основанными на модели Раша [4, 5]. Накопленный за последние десятилетия методический опыт показывает, что в задачах с большим числом одновременных сравнений использование только номинального уровня значимости $p < 0,05$ приводит к завышению доли ложных срабатываний; для контроля этого эффекта рекомендуется процедура контроля ложных

обнаружений (False Discovery Rate, FDR), предложенная Бенджамини и Хохбергом [7]. Современные обзоры методов DIF подтверждают, что сочетание содержательного критерия величины эффекта (например, порогового значения DIF) со статистическим контролем множественных сравнений повышает надежность выводов и снижает риск избыточной интерпретации случайных колебаний как содержательного смещения [7, 8].

Гендерный DIF остается одной из наиболее изучаемых тем прикладной психометрики. Исследования на материале крупномасштабных оценочных программ - от PISA до национальных экзаменов - показывают, что гендерные различия в выполнении отдельных заданий существуют, но носят преимущественно локальный, а не системный характер и заметно зависят от предметной области, формата задания и особенностей выборки [9-13]. Эти результаты подчеркивают необходимость эмпирической проверки DIF для каждого конкретного инструмента, а не экстраполяции выводов, полученных на другом материале.

В казахстанском контексте вопросы психометрического качества и справедливости инструментов оценивания приобретают дополнительную практическую значимость в связи с реформированием системы оценки знаний педагогов и обучающихся [14, 15], а нормативную основу для разработки и применения стандартизированных инструментов оценивания в Республике Казахстан задают соответствующие акты уполномоченного органа в сфере образования [16]. Вместе с тем систематические публикации, в которых DIF-анализ проводился бы одновременно на нескольких параллельных вариантах одного теста для педагогов с явным контролем множественных сравнений, в отечественной литературе представлены ограниченно, что определяет актуальность настоящего исследования.

Объект исследования - результаты выполнения тестовых заданий испытуемыми, различающимися по признаку пола. Предмет исследования - дифференциальное функционирование тестовых заданий (DIF) по признаку пола в трех параллельных вариантах теста для педагогов, рассматриваемое в рамках модели Раша.

Цель исследования - оценить наличие, величину, направление и статистическую устойчивость дифференциального функционирования заданий по признаку пола в трех параллельных вариантах теста для педагогов. Для достижения цели решались следующие задачи: 1) оценить достаточность и структуру выборки по вариантам теста; 2) выявить задания с признаками DIF по критерию DIF Contrast и статистике Rasch-Welch с сопоставлением со статистикой Mantel-Haenszel; 3) применить процедуру контроля ложных обнаружений (FDR) для повышения устойчивости выводов; 4) описать направление выявленных эффектов и сопоставить его с данными других исследований гендерного DIF.

Научная новизна работы состоит в применении комплексного подхода - сочетания критерия величины эффекта, статистики Rasch-Welch, статистики Mantel-Haenszel и FDR-коррекции - к трем параллельным вариантам теста оценки знаний педагогов, используемого в системе оценивания Республики Казахстан. Практическая значимость заключается в том, что результаты могут быть использованы для приоритизации содержательной экспертизы отдельных заданий и совершенствования процедур контроля качества параллельных форм теста.

Материалы и методы исследования

Материалом исследования послужили результаты тестирования, выполненного по трем параллельным вариантам теста для педагогов (условные номера вариантов: 671, 703 и 078), каждый из которых включает 50 дихотомически оцениваемых заданий закрытого типа. Общий объем выборки составил 1417 наблюдений, из них 1071 - женщины (75,6%) и 346 - мужчины (24,4%). Кодировка группы пола в аналитических таблицах Winsteps: D - женский пол, M - мужской пол.

Численность и состав выборки, а также показатели надежности использованных вариантов теста приведены в таблице 1.

Таблица 1 - Характеристика выборки и показатели надежности по трем вариантам теста

Вариант	N	Женщины (D), n	Мужчины (M), n	Женщины, %	Мужчины, %	KR-20	Person reliability
671	302	199	103	65,9	34,1	0,88	0,87
703	488	368	120	75,4	24,6	0,92	0,89
078	627	504	123	80,4	19,6	0,90	0,88
Итого	1417	1071	346	75,6	24,4	0,80-0,92	0,87-0,89
Примечание - составлено авторами на основе данных Winsteps, Table 3.1							

Объем выборок по вариантам составляет от 302 до 627 человек, что в целом соответствует объемам, рекомендуемым для устойчивого DIF-анализа непараметрическими и Rasch-based методами [7]. Вместе с тем численность мужской подгруппы (от 103 до 123 человек по вариантам) заметно уступает численности женской подгруппы, что закономерно отражает гендерный состав контингента педагогов и является типичной особенностью выборок в исследованиях этой профессиональной группы [11]. Такая асимметрия увеличивает стандартную ошибку DIF-оценок для менее многочисленной группы и требует осторожности при интерпретации эффектов, близких к пороговому значению. Об этом подробнее нами будет рассмотрено в разделе «Заключение».

Расчеты выполнены в программе Winsteps 5.8.3.0 в рамках однопараметрической модели Раша [1, 2]. Основным источником данных для DIF-анализа послужила Table 30.1 Winsteps, содержащая для каждого задания групповые оценки трудности (DIF Measure) для женщин (D) и мужчин (M), величину DIF Contrast, ее стандартную ошибку, значение статистики Rasch-Welch t и соответствующий p -value, а также статистику Mantel-Haenszel. Table 3.1 использовалась для оценки объема выборки и показателей надежности теста (KR-20, Person reliability).

Поскольку в выводимых таблицах Winsteps первой указывается группа D, величина DIF Contrast рассчитывалась как разность D - M. Положительное значение DIF Contrast означает, что задание оценено как более трудное для женщин и относительно благоприятнее для мужчин; отрицательное значение означает обратное соотношение.

В качестве первичного критерия выявления DIF использовалось сочетание содержательного и статистического порогов: абсолютная величина $|DIF Contrast| \geq 0,50$ логита при $p < 0,05$ по статистике Rasch-Welch. Пороговое значение 0,50 логита соответствует ориентиру, рекомендуемому руководством Winsteps для заметного DIF-эффекта, и характеризует эффект как минимум умеренной величины [2, 9]. Дополнительно для каждого задания рассматривалось согласие с результатом статистики Mantel-Haenszel [4, 5], что позволяет оценить устойчивость сигнала при использовании методически независимого подхода.

Поскольку в каждом варианте одновременно проверялось 50 заданий, для контроля вероятности ложных обнаружений применялась процедура Бенджамини-Хохберга (FDR) при уровне $q < 0,05$ [6]. FDR рассчитывался отдельно внутри каждого варианта по p -value статистики Rasch-Welch. Такой раздельный расчет учитывает, что варианты представляют собой независимые наборы гипотез.

Наличие статистически значимого и устойчивого после FDR-коррекции DIF интерпретируется в настоящей работе как свидетельство статистической неинвариантности параметра трудности задания между женщинами и мужчинами при контроле общего уровня измеряемой способности. Такой результат не интерпретируется как доказанная содержательная или гендерная предвзятость конкретного задания, поскольку текст задания, его тематическая привязка и когнитивная структура в рамках настоящего анализа не

рассматривались; это ограничение согласуется с общепринятой методической позицией в литературе по DIF [7, 8].

Результаты и обсуждение

Сводные результаты DIF-анализа по трем вариантам теста представлены в таблице 2.

Таблица 2 - Сводные результаты DIF-анализа по трем вариантам теста

Вариант	Заданий	DIF по первичному критерию	Доля, %	Сохранились после FDR
671	50	5	10,0	1
703	50	13	26,0	5
078	50	7	14,0	3
Итого	150	25	16,7	9 (6,0%)

Примечание - составлено авторами на основе данных Winsteps, Table 30.1

По первичному критерию DIF выявлен в 25 из 150 позиций «задание × вариант» (16,7%). Из них 17 позиций одновременно имели $p < 0,05$ по статистике Mantel-Haenszel, что говорит о согласованности результатов двух методически различных подходов для большинства выявленных сигналов. После контроля FDR устойчивыми остались 9 позиций (6,0% от общего числа проверенных заданий). Наибольшая концентрация DIF-сигналов как по первичному критерию, так и после FDR-коррекции наблюдалась в варианте 703 (см. рисунок 1).

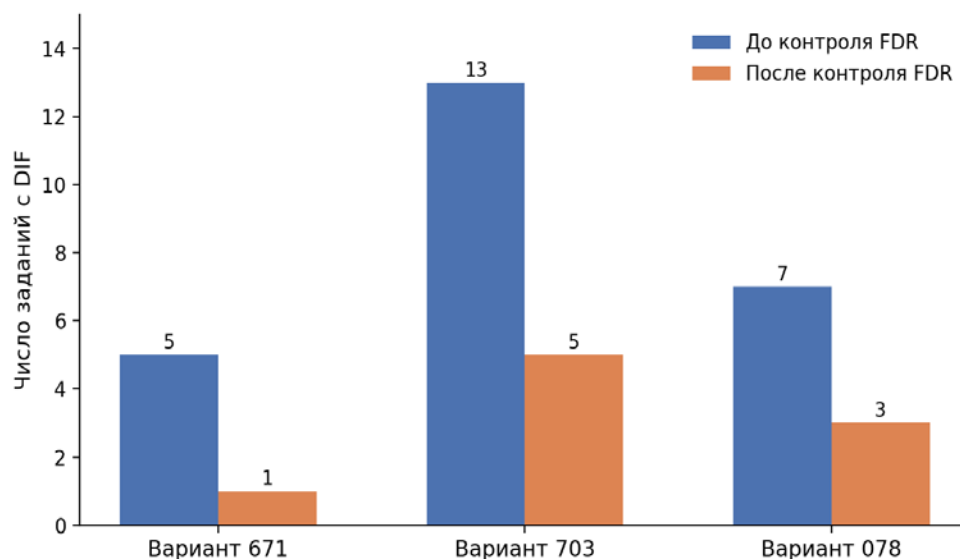


Рисунок 1 - Число заданий с DIF по вариантам теста, до и после контроля FDR

Примечание - составлено авторами на основе данных таблицы 2

Направление DIF-эффектов по вариантам приведено в таблице 3.

Таблица 3 - Направление DIF-эффектов по признаку пола

Вариант	DIF всего	Труднее для женщин (+)	Труднее для мужчин (-)	После FDR
671	5	2	3	1 (0+ / 1-)
703	13	7	6	5 (3+ / 2-)
078	7	4	3	3 (1+ / 2-)
Итого	25	13	12	9 (4+ / 5-)

Примечание - составлено авторами на основе данных таблицы 30.1 Winsteps

При первичном скрининге направление DIF в совокупности трех вариантов оказалось практически сбалансированным: 13 эффектов положительные (задание труднее для женщин) и 12 отрицательных (задание труднее для мужчин). Среди 9 эффектов, сохранившихся после FDR-коррекции, 4 положительные и 5 отрицательных. Распределение отдельных значений DIF Contrast относительно порогового уровня $\pm 0,50$ логита показано на рисунке 2.

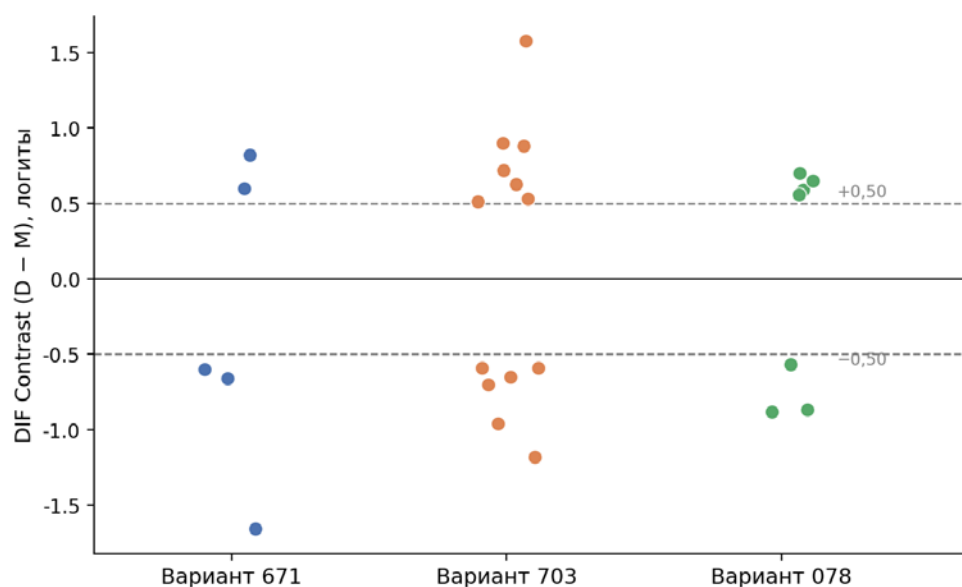


Рисунок 2 - Распределение DIF Contrast по трем вариантам теста
 (пунктирные линии - пороговый уровень $\pm 0,50$ логита; положительные значения соответствуют большей трудности для женщин, отрицательные - для мужчин)
 Примечание - составлено авторами на основе данных таблицы 30.1 Winsteps

Вариант 671.

По первичному критерию в варианте 671 выявлено 5 заданий с DIF (10,0% от общего числа): № 3, 11, 14, 20 и 34. Два из пяти сигналов также статистически значимы по критерию Mantel-Haenszel. После FDR-коррекции сохраняется одно задание - № 14 (таблица 4).

Таблица 4 - Задания с DIF в варианте 671

Задание	DIF женщин (D)	DIF мужчин (M)	DIF Contrast (D-M)	SE	p (Rasch-Welch)	q (FDR)
3	1,32	0,49	0,82	0,29	0,0044	0,1100
11	-0,99	-0,33	-0,66	0,28	0,0169	0,2817
14	1,68	3,34	-1,66	0,46	0,0004	0,0200
20	0,62	0,01	0,60	0,27	0,0276	0,3394
34	0,62	1,21	-0,60	0,29	0,0384	0,3394
Примечание - составлено авторами на основе данных таблицы 30.1 Winsteps						

Наибольшая по модулю величина DIF в этом варианте зафиксирована для задания № 14 (DIF Contrast = -1,66 логита; $p = 0,0004$; сохраняется после FDR при $q = 0,0200$). Отрицательный знак указывает, что данное задание оценено как более трудное для мужчин, чем для женщин.

Вариант 703.

По первичному критерию в варианте 703 выявлено 13 заданий с DIF (26,0%): № 3, 9, 11, 12, 18, 22, 23, 27, 29, 39, 45, 47 и 49 - наибольшее число среди трех рассмотренных вариантов. Девять из тринадцати сигналов также статистически значимы по критерию Mantel-Haenszel. После FDR-коррекции сохраняются задания № 9, 11, 22, 23 и 39 (таблица 5).

Таблица 5 - Задания с DIF в варианте 703

Задание	DIF женщин (D)	DIF мужчин (M)	DIF Contrast (D-M)	SE	p (Rasch-Welch)	q (FDR)
3	-0,45	0,14	-0,59	0,26	0,0229	0,1145
9	-0,18	-1,06	0,88	0,28	0,0022	0,0220
11	4,07	2,49	1,58	0,32	<0,0001	<0,0001
12	-1,05	-0,46	-0,59	0,28	0,0349	0,1442
18	-0,34	-1,06	0,72	0,29	0,0123	0,0689
22	-0,51	0,45	-0,96	0,26	0,0002	0,0025
23	-0,09	1,10	-1,18	0,25	<0,0001	<0,0001
27	0,50	-0,13	0,63	0,25	0,0124	0,0689
29	0,22	-0,32	0,53	0,25	0,0372	0,1442
39	1,43	0,53	0,90	0,24	0,0002	0,0025
45	-0,97	-0,27	-0,70	0,27	0,0106	0,0689
47	0,27	0,92	-0,65	0,25	0,0086	0,0689
49	1,65	1,14	0,51	0,24	0,0375	0,1442
Примечание - составлено авторами на основе данных таблицы 30.1 Winsteps						

Наибольшая по модулю величина DIF в варианте 703 зафиксирована для задания № 11 (DIF Contrast = +1,58 логита; $p < 0,0001$; сохраняется после FDR). Положительный знак указывает, что задание оценено как более трудное для женщин, чем для мужчин. Более

высокая частота DIF-сигналов в этом варианте по сравнению с двумя другими (26,0% против 10,0% и 14,0% по первичному критерию) позволяет рассматривать его как приоритетный объект для последующей содержательной экспертизы заданий, хотя сама по себе более высокая частота DIF не свидетельствует о предвзятости варианта в целом.

Вариант 078.

По первичному критерию в варианте 078 выявлено 7 заданий с DIF (14,0%): № 8, 15, 18, 27, 28, 29 и 34. Шесть из семи сигналов также статистически значимы по критерию Mantel-Haenszel. После FDR-коррекции сохраняются задания № 27, 29 и 34 (таблица 6).

Таблица 6 - Задания с DIF в варианте 078

Задание	DIF женщин (D)	DIF мужчин (M)	DIF Contrast (D-M)	SE	p (Rasch-Welch)	q (FDR)
8	1,39	0,80	0,59	0,23	0,0097	0,0970
15	-0,41	0,16	-0,57	0,24	0,0193	0,1457
18	2,37	1,81	0,56	0,24	0,0204	0,1457
27	-1,03	-0,16	-0,88	0,26	0,0008	0,0350
28	1,37	0,72	0,65	0,23	0,0044	0,0550
29	-1,22	-0,35	-0,87	0,27	0,0014	0,0350
34	0,95	0,25	0,70	0,23	0,0028	0,0467
Примечание - составлено авторами на основе данных таблицы 30.1 Winsteps						

Наибольшая по модулю величина DIF в варианте 078 зафиксирована для задания № 27 (DIF Contrast = -0,88 логита; $p = 0,0008$; сохраняется после FDR). Отрицательный знак указывает, что задание оценено как более трудное для мужчин, чем для женщин.

Обсуждение результатов

Полученная картина в целом согласуется с представлением о гендерном DIF, сложившимся в международной литературе: доля заданий с признаками DIF варьирует от теста к тесту и от варианта к варианту, эффекты чаще носят локальный, а не системный характер, а направление смещения не образует устойчивого перекоса в пользу одной из групп [9-11]. В настоящем исследовании после статистически более строгого критерия (контроль FDR) доля заданий с устойчивым DIF снизилась с 16,7% до 6,0%, что подчеркивает практическую значимость контроля множественных сравнений при анализе тестов с большим числом заданий [6, 7].

Различия между вариантами по частоте DIF - от 10,0% в варианте 671 до 26,0% в варианте 703 по первичному критерию - могут объясняться как особенностями конкретных заданий, так и структурой выборки (например, распределением уровня способности внутри гендерных подгрупп для данного варианта). Поскольку тематическое содержание заданий не анализировалось, в рамках настоящей работы невозможно установить содержательную причину более высокой частоты DIF-сигналов в варианте 703. Это направление обозначено как перспектива дальнейших исследований.

Сопоставление с работами, изучавшими источники гендерного DIF через анализ дистракторов и форматов заданий [10-12], указывает на то, что дальнейшее уточнение природы выявленных эффектов потребует привлечения содержательной экспертизы конкретных заданий, привязки к предметной области и, при необходимости, анализа типа

ответа. Такой шаг выходит за рамки настоящего психометрического анализа, но представляется логичным продолжением работы.

С практической точки зрения результаты имеют значение для процедур контроля качества параллельных форм теста в РГП на ПХВ «Национальный центр тестирования»: регулярный DIF-анализ по полу с обязательным контролем множественных сравнений позволяет заблаговременно выявлять задания, заслуживающие дополнительной содержательной экспертизы.

Заключение

Проведенный психометрический анализ трех параллельных вариантов теста для педагогов (671, 703, 078; $N = 1417$) в рамках модели Раша показал, что по первичному критерию ($|DIF\ Contrast| \geq 0,50$ логита, $p < 0,05$) дифференциальное функционирование по признаку пола выявляется в 25 из 150 позиций «задание $\times$ вариант» (16,7%), а после контроля ложных обнаружений (FDR) статистически устойчивым остается существенно меньшее число сигналов - 9 позиций (6,0%). Гипотеза о том, что выявленные эффекты будут носить локальный, а не системный характер, в целом подтвердилась: направление устойчивых после FDR эффектов оказалось практически сбалансированным (4 задания труднее для женщин, 5 - для мужчин), что не дает оснований говорить о системном смещении трудности использованных вариантов теста в пользу одного пола.

Полученные результаты имеют значение как для теории, так и для практики педагогических измерений. В теоретическом плане они подтверждают целесообразность сочетания содержательного порога величины эффекта, независимой проверки по статистике Mantel-Haenszel и статистического контроля множественных сравнений (FDR) при анализе DIF на тестах с большим числом одновременно проверяемых заданий. В практическом плане результаты позволяют рекомендовать: 1) приоритетную содержательную экспертизу заданий, сохранивших статистическую значимость DIF после FDR-коррекции, в первую очередь в варианте 703, где концентрация сигналов оказалась наибольшей; 2) включение DIF-анализа с контролем множественных сравнений в стандартную процедуру психометрической проверки параллельных форм теста; 3) регулярный мониторинг DIF по полу при последующих циклах тестирования педагогов для отслеживания устойчивости выявленных закономерностей во времени.

Ограничения исследования связаны с тем, что: содержание и тематическая привязка заданий не анализировались, в связи с чем выявленный DIF не может интерпретироваться как доказанная содержательная предвзятость конкретных заданий; численность мужской подгруппы заметно меньше численности женской подгруппы во всех трех вариантах, что увеличивает стандартные ошибки DIF-оценок и требует осторожной интерпретации эффектов, близких к пороговому значению; анализ ограничен тремя вариантами теста и не включает проверку устойчивости результатов во времени. Перспективным направлением дальнейших исследований представляется содержательный анализ заданий с устойчивым DIF, а также расширение выборки вариантов и циклов тестирования для повторной проверки выявленных закономерностей.

Информация о финансировании

Работа выполнена в рамках гранта внутреннего финансирования РГП на ПХВ «Национальный центр тестирования» МНВО РК «Психометрическое исследование качества тестовых заданий системы оценки знаний педагогов (ОЗП) и разработка рекомендаций по их совершенствованию».

Список литературы

1. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. - Chicago : The University of Chicago Press, 1980 (переизд.). - 199 p.
2. Linacre J.M. Winsteps® Rasch Measurement Computer Program User's Guide. - Portland, Oregon: Winsteps.com, 2024. - <https://www.winsteps.com/winman/index.htm?differentialitemfunctioningdif.htm>
3. Hambleton R.K., Swaminathan H., Rogers H.J. Fundamentals of Item Response Theory. - Newbury Park, CA : Sage Publications, 1991. - 174 p.
4. Camilli G., Shepard L.A. Methods for Identifying Biased Test Items. - Thousand Oaks, CA : Sage Publications, 1994. - 174 p.
5. Holland P.W., Thayer D.T. Differential item performance and the Mantel-Haenszel procedure // Test Validity / eds. H. Wainer, H.I. Braun. - Hillsdale, NJ : Lawrence Erlbaum Associates, 1988. - P. 129-145.
6. Mantel N., Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease // Journal of the National Cancer Institute. - 1959. - Vol. 22, № 4. - P. 719-748.
7. Benjamini Y., Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing // Journal of the Royal Statistical Society: Series B (Methodological). - 1995. - Vol. 57, № 1. - P. 289-300. - DOI: 10.2307/2346101
8. Mapuranga R., Dorans N.J., Middleton K. A Review of Recent Developments in Differential Item Functioning : ETS Research Report RR-08-43. - Princeton, NJ : Educational Testing Service, 2008. - 32 p. - <https://doi.org/10.1002/j.2333-8504.2008.tb02129.x>
9. Monahan P.O., McHorney C.A., Stump T.E., Perkins A.J. Odds ratio, delta, ETS classification, and standardization measures of DIF magnitude for binary logistic regression // Journal of Educational and Behavioral Statistics. - 2007. - Vol. 32, № 1. - P. 92-109. - <https://doi.org/10.3102/1076998606298035>
10. Liu Y., Yin H., Xin T., Shao L., Yuan L. A comparison of differential item functioning detection methods in cognitive diagnostic models // Frontiers in Psychology. - 2019. - Vol. 10. - Article 1137. - <https://doi.org/10.3389/fpsyg.2019.01137>
11. Liu O.L., Wilson M. Gender differences in large-scale math assessments: PISA trend 2000 and 2003 // Applied Measurement in Education. - 2009. - Vol. 22, № 2. - P. 164-184. - <https://doi.org/10.1080/08957340902754635>
12. Susongko P., Arfiani Y., Kusuma M. Determination of gender differential item functioning in Tegal-students' scientific literacy skills with integrated science (SLiSIS) test using Rasch model // Jurnal Pendidikan IPA Indonesia. - 2021. - Vol. 10, № 2. - P. 270-281. - <https://doi.org/10.15294/jpii.v10i2.26775>
13. Ozdemir B., AlGhamdi H.M. Investigating the distractors to explain DIF effects across gender in large-scale tests with non-linear logistic regression models // Frontiers in Education. - 2022. - Vol. 6. - Article 748884. - <https://doi.org/10.3389/feduc.2021.748884>
14. Shear B.R. Gender bias in test item formats: Evidence from PISA 2009, 2012, and 2015 math and reading tests // Journal of Educational Measurement. - 2020. - Vol. 57, № 3. - P. 369-396. - <https://doi.org/10.1111/jedm.12372>
15. Карданова Е.Ю., Иванова А.Е. Психометрические исследования: современные методы и новые возможности для образования // Вопросы образования / Educational Studies Moscow. - 2023. - № 3. - С. 8-19. - <https://doi.org/10.17323/vo-2023-17951>
16. Абдрашилов Б.С., Алтыбаева Ш.Б., Шинетова Л.Е. Результаты исследования системы ЕНТ, проведенного в рамках проекта Всемирного банка // Педагогические измерения. - 2025. - № 1(1). - С. 5-19. - <https://doi.org/10.63597/УТО3105-4161.2025.1.1.001>

Ж.Ж. Бейсенова, Л.Е. Шинетова, Г.К. Жабаетова, С.Ж. Анафияева

**РАШ МОДЕЛІ НЕГІЗІНДЕ ПЕДАГОГТАРДЫ ТЕСТІЛЕУДЕГІ ТЕСТ
ТАПСЫРМАЛАРЫНЫҢ ДИФФЕРЕНЦИАЛДЫ ЖҰМЫС ІСТЕУІ**

Мақалада педагогтерге арналған тесттің үш параллель нұсқасының материалдары негізінде жыныс белгісі бойынша тест тапсырмаларының дифференциалды жұмыс істеуін (DIF) жүргізілген психометрикалық талдау нәтижелері ұсынылған. Зерттеудің мақсаты – Раш моделі аясында DIF әсерлерінің болуын, шамасы мен бағытын бағалау, сондай-ақ көптік салыстыруларды бақылау кезінде олардың статистикалық тұрақтылығын анықтау. Параллель нұсқалардың қиындық деңгейлері

өзара салыстырмалы болған жағдайда DIF әсерлерінің үлесі мен бағыты жекелеген тапсырмалармен шектеліп, белгілі бір жыныс тобының пайдасына жүйелі ығысуды қалыптастырмайды деген болжам жасалды. Зерттеу материалы ретінде 1417 тестіленушінің (1071 әйел және 346 ер адам) тестілеу нәтижелері пайдаланылды. Тестіленушілер әрқайсысы дихотомиялық тәсілмен бағаланатын 50 тапсырмадан тұратын тесттің үш нұсқасын орындады. Деректерді өңдеу Winsteps 5.8.3.0 бағдарламасында жүргізілді. Тапсырмалардың дифференциалды жұмысы DIF шамасы бойынша бағаланды және қосымша Mantel-Haenszel статистикасы арқылы тексерілді. Жалған анықтаулар үлесін (FDR) бақылау үшін әрбір тест нұсқасына жеке қолданылған Бенджамини-Хохберг процедурасы пайдаланылды. DIF-тің бастапқы критерийі бойынша «тапсырма × нұсқа» түріндегі 150 позицияның 25-інде (16,7%) DIF анықталды. FDR түзетуінен кейін статистикалық тұрғыдан тұрақты 9 позиция (6,0%) қалды. Тұрақты әсерлердің бағыты іс жүзінде теңгерімді болды: 4 тапсырма әйелдер үшін, ал 5 тапсырма ер адамдар үшін қиынырақ болды. Бұл тест тапсырмалары қиындықтарының белгілі бір жыныс тобының пайдасына жүйелі түрде ығысқанын айтуға негіз бермейді. DIF-сигналдарының ең жоғары шоғырлануы 703-нұсқада байқалды. Жалпы айтқанда, алынған нәтижелер пайдаланылған тест нұсқаларының жыныс белгісі бойынша өлшеу инварианттылығының қолайлы деңгейін көрсетеді және тесттің параллель нұсқаларын әзірлеу мен қолдану барысында DIF-ке тұрақты мониторинг жүргізу жөнінде ұсынымдар әзірлеуге мүмкіндік береді.

Түйін сөздер: тапсырмалардың дифференциалды жұмыс істеуі (DIF), Раш моделі, өлшеу инварианттылығы, педагогтарды тестілеу, гендерлік айырмашылықтар, Mantel-Haenszel статистикасы.

Zh.Zh. Beisenova, L.E. Shinetova, G.K. Zhabayeva, S.Zh. Anafiyeva

DIFFERENTIAL ITEM FUNCTIONING IN TEACHER TESTING BASED ON THE RASCH MODEL

This article presents the results of a psychometric analysis of differential item functioning (DIF) by gender based on three parallel forms of a test for teachers. The aim of the study was to assess the presence, magnitude, and direction of DIF effects within the Rasch model and to determine their statistical robustness while controlling for multiple comparisons. It was hypothesized that, given comparable difficulty across the parallel test forms, the proportion and direction of DIF effects would be limited to individual items and would not indicate systematic bias in favor of either gender. The study was based on the test results of 1,417 examinees (1,071 women and 346 men) who completed three test forms, each consisting of 50 dichotomously scored items. Data were analyzed using Winsteps 5.8.3.0. Differential item functioning was assessed based on the magnitude of DIF, with additional testing using the Mantel–Haenszel statistic and control of the false discovery rate (FDR) using the Benjamini–Hochberg procedure, applied separately to each test form. According to the initial DIF criterion, DIF was identified in 25 of the 150 item × form positions (16.7%). After FDR correction, 9 positions (6.0%) remained statistically robust. The direction of the robust effects was nearly balanced: 4 items were more difficult for women and 5 items were more difficult for men, providing no evidence of a systematic shift in item difficulty in favor of either gender. The highest concentration of DIF signals was observed in Form 703. Overall, the results indicate an acceptable level of measurement invariance across gender for the test forms examined and provide a basis for recommendations regarding regular DIF monitoring in the development and operational use of parallel test forms.

Keywords: differential item functioning (DIF), Rasch model, measurement invariance, teacher testing, gender differences, Mantel–Haenszel statistic.

References

1. Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago, IL: The University of Chicago Press.
2. Linacre, J. M. (2024). Winsteps® Rasch measurement computer program user's guide. Table 30.1 "Differential item functioning (DIF) pairwise." Portland, OR: Winsteps.com. Retrieved from <https://www.winsteps.com/winman/index.htm?differentialitemfunctioningdif.htm>

3. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage Publications.
4. Camilli, G., & Shepard, L. A. (1994). Methods for identifying biased test items. Thousand Oaks, CA: Sage Publications.
5. Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.
6. Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22(4), 719–748.
7. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
8. Mapuranga, R., Dorans, N. J., & Middleton, K. (2008). A review of recent developments in differential item functioning (ETS Research Report RR-08-43). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2008.tb02129.x>
9. Monahan, P. O., McHorney, C. A., Stump, T. E., & Perkins, A. J. (2007). Odds ratio, delta, ETS classification, and standardization measures of DIF magnitude for binary logistic regression. Journal of Educational and Behavioral Statistics, 32(1), 92–109. <https://doi.org/10.3102/1076998606298035>
10. Liu, Y., Yin, H., Xin, T., Shao, L., & Yuan, L. (2019). A comparison of differential item functioning detection methods in cognitive diagnostic models. Frontiers in Psychology, 10, Article 1137. <https://doi.org/10.3389/fpsyg.2019.01137>
11. Liu, O. L., & Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. Applied Measurement in Education, 22(2), 164–184. <https://doi.org/10.1080/08957340902754635>
12. Susongko, P., Arfiani, Y., & Kusuma, M. (2021). Determination of gender differential item functioning in Tegal-students' scientific literacy skills with integrated science (SLiSIS) test using Rasch model. Jurnal Pendidikan IPA Indonesia, 10(2), 270–281. <https://doi.org/10.15294/jpii.v10i2.26775>
13. Ozdemir, B., & AlGhamdi, H. M. (2022). Investigating the distractors to explain DIF effects across gender in large-scale tests with non-linear logistic regression models. Frontiers in Education, 6, Article 748884. <https://doi.org/10.3389/educ.2021.748884>
14. Shear, B. R. (2020). Gender bias in test item formats: Evidence from PISA 2009, 2012, and 2015 math and reading tests. Journal of Educational Measurement, 57(3), 369–396. <https://doi.org/10.1111/jedm.12372>
15. Kardanova, E. Y., & Ivanova, A. E. (2023). Psikhometricheskie issledovaniya: sovremennyye metody i novye vozmozhnosti dlya obrazovaniya [Psychometric studies: Modern methods and new opportunities for education]. Voprosy obrazovaniya / Educational Studies Moscow, (3), 8–19. <https://doi.org/10.17323/vo-2023-17951>
16. Abdrasilov, B. S., Altybayeva, Sh. B., & Shinetova, L. E. (2025). Rezul'taty issledovaniya sistemy ENT, provedennogo v ramkakh proekta Vsemirnogo banka [Results of the study of the UNT system conducted within the World Bank project]. Pedagogikalyq ölsheuder / Pedagogicheskie izmereniya / Pedagogical Measurements, (1), 6. <https://doi.org/10.63597/UTO3105-4161.2025.1.1.001>

Сведения об авторах:

Бейсенова Жанар Жумагазиновна (автор-корреспондент) - магистр, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, руководитель отдела науки и развития рынка, Астана, Республика Казахстан, e-mail: zhanar-beisenova@mail.ru

Шинетова Ляззат Ермаковна – магистр, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, исполняющий обязанности заместителя директора, Астана, Республика Казахстан, e-mail: shinetovalyazzat24@gmail.com

Жабасева Гульдана Куатовна – магистр, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, ведущий научный сотрудник отдела науки и развития рынка, Астана, Республика Казахстан, e-mail: guldanzhabayeva@gmail.com

Анафияева Сулупан Жанабаевна – магистр, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, ведущий научный сотрудник отдела психометрического анализа, Астана, Республика Казахстан, e-mail: sulupaaan@gmail.com

Авторлар туралы мәлімет:

Бейсенова Жанар Жумагазиновна (автор-корреспондент) - магистр, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, ғылым және нарықты дамыту бөлімінің басшысы, Астана, Қазақстан Республикасы, e-mail: zhanar-beisenova@mail.ru

Шинетова Ляззат Ермековна – магистр, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, директор орынбасарының міндетін атқарушы, Астана, Қазақстан, e-mail: shinetovalyazzat24@gmail.com

Жабаетова Гүлдана Қуатовна – магистр, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Ғылым мен нарықтық дамыту бөлімінің жетекші ғылыми қызметкері, Астана, Қазақстан Республикасы, e-mail: guldanzhabayeva@gmail.com

Анафияева Сулупан Жанабаетовна – магистр, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Психометрикалық талдау бөлімінің жетекші ғылыми қызметкері, Астана, Қазақстан Республикасы, e-mail: sulupaaan@gmail.com

Information about authors:

Beisenova Zhanar Zhumagazinovna (corresponding author) – Master’s degree, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Head of the Department of Science and Market Development, Astana, Republic of Kazakhstan, e-mail: zhanar-beisenova@mail.ru

Shinetovalyazzat Ermekovna – Master’s degree, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Acting Deputy Director, Astana, Republic of Kazakhstan, e-mail: shinetovalyazzat24@gmail.com

Zhabayeva Guldana Kuvatovna – Master’s degree, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Leading Researcher, Department of Science and Market Development, Astana, Republic of Kazakhstan, e-mail: guldanzhabayeva@gmail.com

Anafiyeva Sulupan Zhanabayevna – Master’s degree, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Leading Researcher, Department of Psychometric Analysis, Astana, Republic of Kazakhstan, e-mail: sulupaaan@gmail.com

L. Rizayeva¹, A. Zhunisova^{2*}, L. Nurseitova³, K. Turalbaeva⁴, L. Seisekenova⁵

^{1,2,3}Zh.A.Tashenev University, Shymkent, Kazakhstan

⁴Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan

⁵Kazakh National Pedagogical University named after Abai, Aimagy, Kazakhstan

*e-mail: aknura_er74@mail.ru

¹ORCID 0000-0002-5924-9973, ²ORCID 0009-0000-9338-1321, ³ORCID 0009-0004-3687-6140,

⁴ORCID 0009-0004-6640-0522, ⁵ORCID 0009-0006-9189-5410

DIAGNOSTICS OF THE FORMATION OF RESEARCH SKILLS OF PRIMARY SCHOOL CHILDREN IN MATEMATICS LESSONS

The article discusses the problem of the formation of research skills of Primary school students in mathematics lessons. The relevance of the study is determined based on regulatory documents. In the course of the study, the essence of the concept of research skills in domestic and foreign studies through the theoretical method and the advantages of research skills are described. In accordance with the motivational, substantive and reflexive components of the formation of research skills of primary school students are selected through diagnostic methods. The results of the empirical study made it possible to comprehensively diagnose the level of development of students' research skills for each component. According to the obtained quantitative data, it was concluded that a formative research stage is necessary due to the low level of research skills of primary school students. The results of the study provide the following scientifically based recommendations: systematically introduce problem situations, logical tasks, and tasks of a research nature into the content of primary school mathematics lessons in order to develop students; develop a system of methodological tools and tasks for primary school teachers to organize students' research work. This study offers practical implications for the preparation and continuing professional development of primary school teachers. Diagnostic data on the development of research skills among primary school students in mathematics lessons may be incorporated into the educational process of pedagogical universities and colleges. In addition, the findings can serve as a resource for institutions providing teacher preparation, retraining, and professional development programmes.

Keywords: research skills, mathematics, interest, design, logical thinking, creativity, diagnostics.

Introduction

The need for the formation of a fully developed creative and research personality is growing. One of the reasons for this is the rapid development of modern science and technology, in an era of increased data flow. In order to quickly adapt to such innovative changes, it is necessary to master the level of search, research, and logical thinking within the framework of primary education subjects.

In this regard, the Law of the Republic of Kazakhstan “on education” obliges to increase the competitiveness of individuals who are concerned about the fate of the country of tomorrow [1].

“State General Obligatory Standards of preschool education, and training, primary, basic secondary, general secondary, technical and vocational, postsecondary education” [2] and the curriculum of grades 1-4 of the academic discipline “Mathematics” academic subject [3] define the main goals and objectives of Education, emphasize the need to creatively apply knowledge, think critically and, most importantly, form students skills’ to conduct research.

Additionally, the concept of development of preschool, secondary, technical and vocational education in the Republic of Kazakhstan for 2023-2029 reflects the development of functional literacy, critical thinking and research skills of students as priorities of the education system [4].

The announcement by the head of State K. Tokayev at the second central Asia-China summit next year as the “Year of Education and Scientific Research” [5] indicates the need to pay special

attention to the research work of students in the future and pay special attention to the formation of their research skills.

And this increases the need for primary school students not only to acquire knowledge, but also to gain comprehensive experience, to independently set learning goals, to design ways to implement them, to develop skills to monitor and evaluate their achievements, especially in engaging in research activities. In this regard, the problem of forming students' research skills by solving problems in elementary school mathematics education is highly crucial.

Theoretical studies indicate that certain aspects of fostering research skills in primary school students have been examined. However, the issue of developing analytical skills in students during mathematics instruction in primary grades remains insufficiently explored. This highlights the relevance of addressing the development of research skills among primary school students in mathematics lessons.

The purpose of the study is to diagnose the formation of research skills of primary school children in mathematics lessons.

Research objectives: to determine the essence of the concept of “research skill”; to determine the components, criteria and indicators of the formation of analytical skills in primary school children in mathematics lessons, to obtain the initial data on the formation of primary school children in mathematics lessons.

The scientific novelty and practical significance of the work are demonstrated by the diagnostic assessment of the formation of primary school students' research skills during mathematics lessons, as well as by the applicability of the research findings in higher pedagogical institutions and colleges, and in institutes for the training, retraining, and professional development of primary school teachers.

Materials and research methods

The study was grounded in a theoretical approach that enabled the clarification of key concepts related to the research topic, an examination of the the importance of research skills development among primary school students, and a review of relevant scientific literature. Empirical methods were subsequently applied to examine the research problem in practice, with the collected data presented in quantitative form. Specifically, the following were used: G.N. Kazantseva's “Studying Attitudes to the Subject of Research” method, the “Questionnaire for Determining Activity and Motivation for Learning Mathematics” (Author's Questionnaire), the “Determining Students' Research Skills” diagnostic chart (based on A.I. Savenkov), the “My Attitude to Research Activity” questionnaire (modified by Yu.A. Kazimirova), and A.V. Karpov's “Diagnosing the Level of Reflexivity Development” method. These methods were selected taking into account the motivational, substantive, and reflexive components of developing students' research skills in mathematics.

The following data were obtained as a result of a theoretical study of the study:

The concept of “Research” can be viewed as “both scientific work and a type of human activity aimed at obtaining knowledge, understanding and evaluating the essence, specificity, and patterns of development of phenomena, as well as searching for opportunities to apply the acquired knowledge, including in the study of the natural environment” [6].

Research skills are a person's ability to create new concepts and ideas by collecting data. These skills include methods, documentation, and interpretation of collected data. The research is conducted to evaluate hypotheses and disseminate the results in the most appropriate ways. Research skills are involved with experience [6].

Studies indexed in the Scopus database also addresses the issue of developing research skills among primary school students. A group of scholars led by Rizayeva examined the potential of mathematical problems in fostering research skills [7], while Fry, Nakar, and Zorn investigated the possibilities of an inquiry-based approach in mathematics and the pedagogical challenges associated with its implementation [8]. Domestic studies have explored the theoretical and diagnostic aspects

of cultivating a research culture and research skills [9, 10]. While Bapanova and her colleagues emphasized the need to employ diagnostic methods and criteria in assessing research skills [11], López Chira empirically examined an educational model aimed at developing the research skills of primary school students [12]. These studies form the theoretical basis of our research.

Sky Ariella defines research skills as “the ability to find specific information on a certain topic” [13]. He says that research skills include “the ability to identify the necessary data, find and formulate the results, and then explain to others”. The author offers 12 much more important research skills, which include: data collection; analysing information and data from various sources; searching for information on the internet; writing reports; getting the interviews; planning and charting; critical thinking; studying the necessary issues; setting time; communication skills; using your own network; managing; assigning meaning, despite all the big and small information [13].

Advantages of research skills:

- increases interest: interest – curiosity is the passion for learning new things and a powerful engine of learning. Inquisitive people are able to listen well and are willing to consider not only their own ideas and perspectives, but also to the ideas and views of other people. Intrigued researchers ask many questions that naturally need to be answered and they especially focus on finding answers to the questions.

- Develops self – awareness: in addition to knowing the subjective opinions of others, it is crucial to develop research skills, as well as to know the benefits of knowing what research provides, we get to know many people through our research. From the very beginning of the research, the benefits we derive from it are reflected in beliefs and attitudes and encourage them to open their minds to look at others points of view and things.

- Empowering researchers to share their findings.

Providing researchers with a space and platform to present their research and share their results can be crucial to maintain motivation for future research.

Z.A. Isaeva characterised: “research skills are primarily gnostic and design skills, on the basis of which one can acquire new knowledge and master new actions”, “teacher's research skills are the ability to perform professional teaching activities in new conditions and new approaches and tools, the ability to apply research methods” [14]. We believe that, agreeing with the author's opinion, the solution of educational research tasks for the formation of research skills of primary school students is implemented within the framework of mathematical and logical thinking, the use of research methods and techniques.

Studying the above considerations, we formulate a research skill as an automated skill of students' intellectual and practical abilities necessary for searching, summarising, analysing and evaluating, demonstrating the results of information.

In accordance with the specifics of the mathematics subject, the main focus is on the deep assimilation of knowledge by students. Teachers face obstacles in making effective use of their time for research activities. It also depends on the composition of the research assignments. The study by K.T. Turalbayeva, G.S. Dzharasova, I.Zh. Zhumatova, G.T. Sagnaeva, and B.K. Asemova notes that the development of a database of math assignments is a multi-step process. It includes the stages from the preparation of tasks to verification, correction and approbation [15]. Assignments for developing research skills are also such a multifaceted process in which it is necessary to take into account such factors as the student's capabilities, specifics, condition, environment, etc.

Having studied the analysis that is defined as “the development of students' research skill through solving problems in teaching elementary school mathematics is a thought process aimed at thinking, reasoning, comparing, grouping, mastering, searching methods through solving educational and cognitive research tasks and creative, logical tasks of students”. Thus, an experiment is conducted to achieve the task of determining the level of developing analytical skills that are of great importance in the formation of schoolchildren' intelligence. in mathematics lessons.

The purpose of the ascertaining experiment is to determine the initial level of formation of research skills of primary school students and the process of formation of students' research skills.

Two groups (control and experiment) were taken to carry out experimental work. 63 students of the 3rd grade of the E. Izatullaev Comprehensive school in Turkestan region was acquired which was divided into control (31) and experimental groups (32).

During the organization of the exploratory and experimental work:

- the purpose of pedagogical experimental work is defined;
- the requirements for the experiment and the stages of its implementation are determined;
- the content of the experiment is characterised;
- the composition of the control and experimental groups participating in the experiment is described;
- the methods and approaches of determining the level of developing children's research skills.

To determine the initial level of research skill development, the participants completed a set of diagnostic tasks.

Diagnostic methods selected in accordance with the components of research skill development in primary school mathematics lessons are presented in Table 1.

Table 1 – Methods selected by components

Component	Methods
Motivational	1) Methodology of G. N. Kazantseva “study of attitude to the subject of study” 2) “questionnaire for determining educational activity and motivation in Mathematics” (author's questionnaire)
Content	1) diagnostic card “determining the research skills of students” (according to A. I. Savenkov) 2) questionnaire “my attitude to research activities” (Modification of Yu. a. Kazimirova)
Reflective	Methodology of A.V. Karpov “diagnostics of the level of development of reflexivity”

First, to determine the motivational component, we used the “methodology for studying the attitude of G.N. Kazantseva to the academic subject”. This methodology is based on the DC section, in which we conducted research based on the materials of the first 3 sections.

Section I is devoted to writing the subjects that the student likes the most, as well as the subjects that he likes the most.

Part I:

The best of the subjects studied in school:

A. Subjects that I like: _____

B. subjects that I don't like: _____

The students “I love mathematics” were selected from two questions. Of the two groups that participated in the experiment, 56 out of 63 students said “love”, and 7 wrote “dislike”. The students who answered the question of part I positively were asked questions of part II, “I don't like mathematics” - of part III. In fact, there are 21 questions in both sections, we chose only 11 based on the age characteristics of younger students and depending on our research problem.

Part II.

Express your point of view, complete the sentence:

I love mathematics because...

1. It's an interesting subject.
2. I like that the teacher drives well.
3. Everyone needs mathematics.
4. Easy to learn.
5. Teaches you to think.
6. I'm learning observation and intelligence.

7. My friends love.
8. My parents consider math to be very important.
9. The teacher loves me.
10. It changes the environment.
11. I'm interested .

I don't like math because...

1. I'm bored.
2. I don't like what the teacher gives.
3. Not everyone needs to know this.
4. It's hard to learn.
5. It doesn't teach you to think.
6. there is no need for observation, intelligence.
7. Friends don't like this subject.
8. My parents consider math less important.
9. The teacher hates me.
10. It does not affect the environment.
11. I'm bored.

Based on the results of the conducted methodology, the main, main motives of students in section I, i.e. the love of mathematics:

- It can be mastered very quickly;
- The teacher teaches mathematics in a clear and interesting way;
- Mathematics is a subject that everyone needs, anywhere.

Part II, that is, the motivation of some students for mathematics based on negative reviews:

- Not everyone needs to know math;
- Unclear, difficult;
- Does not understand some tasks.

The following types of educational motives of the attitude of students of grade 3 to the subject were identified; readiness to study mathematics; desire to know mathematics; cognitive and educational motives; understanding the need for mathematics; presence of research activity. Accepted, fulfilling their motivations.

The purpose of the questionnaire for determining educational activity and motivation in mathematics (author's questionnaire) is to determine the educational activity and motivation of Primary School students in mathematics. The questionnaire includes 6 questions. 63 students took part in the survey.

1. Do you enjoy solving math problems?
2. Do you like to solve problems in math class with interest and do additional assignments besides the textbook?
3. Do you have any problems doing your math homework?
4. Do you like solving problems in the textbook and completing tasks in math?
5. Do you enjoy writing scientific projects and doing research?
6. Do you want to have interesting logical research – problems and tasks in maths that develop research skills?

These diagnostic papers were selected based on the motivational component as it allows to identify students' interests and engagement to the subject, activity in mathematics, and attitude to research assignments.

The main goal of this methodology is to determine the initial level of formation of research skills in primary school children. Filling out the card is carried out according to the following criteria:

Low – level children are characterized by an unstable interest in educational and research work, and they can only complete tasks under the strict guidance of a teacher.; middle-level students are determined by their constant interest in educational and research work, the skills to

solve problems with the help of a teacher and the acquisition of individual skills that allow them to conduct educational and scientific research, the expression of creative components in choosing a topic, methods, and presentation of the results of cognitive search.; At a high level, students are characterized by the dominance of internal cognitive motives for educational and research work, the mastery of a set of skills that allow them to independently seek new knowledge, the manifestation of originality at each stage of research, as well as the presence of creativity and curiosity in research activities.

The diagnostic map for determining students' research skills is presented in the table below (Table 2).

Table 2 – Research skills of primary school children

Students names	Research skills of primary school children								total
	finding a problem	ask a question	Proposing a hypothesis	Defining a concept	classifying	observing	the ability to experiment	drawing conclusions	

Assessment criterias:

3 points – skill is formed;

2 балл – skill is being formed;

1 балл – skill is not formed.

The levels of development of research skills were assessed by the following indicators:

24-21 points – high level;

20-16 балл – medium level;

15-10 балл – low level.

The fourth method used is the questionnaire “my attitude to research activities”, adapted by Yu.a. Kazimirova in Kazakh (modified).

In order to determine the attitude of students to research activities and research skills, we changed the questionnaire of Kazimirova Yu.a. The questionnaire includes 4 questions that allow you to fully assess the attitude of children to research work.

1. “Do you want to engage in research activities?”

a) yes, I would like to try;

b) yes, I would like to be very engaged;

c) I have never thought about it;

d) no it does not interest me at all;.

2. “What questions do you like to ask most often?”

a) how does it do the following?

b) why does this happen?

c) how did it come about?

d) I do not like asking questions.

3. “In what extracurricular subject do you like to attend a practical lesson?”

a) I like to go to the circle of artistic creativity, where I like to mix paints and experiment with different materials;

b) I like to go to the math club and solve logical, creative, research problems;

c) I go to the science club, where I like to look at different things under a microscope;

d) I do not participate in practical classes.

4. “Do you want adults to support your interest in scientific research?”

- a) Yes, because it is very interesting
- b) Yes, because I want to learn a lot
- c) Yes, because it will be useful for me in high school
- d) I do not need it, I do not want it

The criteria for evaluating the questionnaire are as follows:

1 question (1,2 – 2 points, 3 – 1 point, 4 – 0 point)

2 questions (1,2,3 – 2 points, 4 – 0 point)

3 questions (1,2,3 – 2 points, 4 – 0 point)

4 questions (1,2,3 – 2 points, 4 – 0 point)

The levels are determined as follows:

0- to 1 point – low level

2- to 4 points – medium, level

5- to 8 points – high level

Using the “Methodology for diagnosing the level of development of reflexivity” by A.V. Karpov, the last of the diagnostic methods - the reflective component of the development of students' research skills through solving problems in teaching elementary mathematics - was identified. Reflection is a single psychological reality both in the cognition of one's own inner world and in the cognition of the external world (including the subjective - “the world of others”). Furthermore, the author conditionally distinguishes two types of reflection: reflection on the content of one's own consciousness and reflection on the course and outcomes of one's own actions, as well as the inner world of another person. Students are asked to respond 27 questions without spending too much time considering their answers and to assign each question a number corresponding to one of the following response options:

- 1 – completely incorrect;
- 2 – incorrect;
- 3 – mostly incorrect;
- 4 – do not know;
- 5 – mostly correct;
- 6 – correct;
- 7 – completely correct.

Among the 27 statements, 15 are formulated as direct items (question numbers: 1, 3, 4, 5, 9, 10, 11, 14, 15, 18, 19, 20, 22, 24, 25). The remaining 12 are inverse statements corresponding to the subjects' answers to direct questions to obtain the final score, and represent inverse numbers that must be taken into account when processing the results if they are replaced by numbers obtained by inverting the response scale, i.e. 1=7, 2=6, 3=5, 4=4, 5=3, 6=2, 7=1. Their numerically replaced indicators are given in Table 3.

Table 3 – Conversion of test scores into Sten values

Sten scores	0	1	2	3	4	5	6	7	8	9	10
test scores	more than 80	81-100	101-107	108-113	114-122	123-130	131-139	140-147	148-156	157-171	more than 172

Reflection is the process of a person's awareness of his inner world (thoughts, feelings, intentions) and the external world (including the inner world of other people). The author divides reflection into two main types;

1. Reflection on the content of one's own consciousness: reflection on the course and results of one's thoughts, feelings, and actions.

2. Reflection of another person's inner world: an attempt to understand other people's thoughts and feelings.

Evaluation of the results according to A.V. Karpov's methodology:

The results are evaluated using A.V. Karpov's methodology. A score of 7 or higher indicates a high level of reflexivity; a score between 4 and 7 signifies an average level; and a score below 4 indicates a low level of reflexivity. The quantitative and qualitative data obtained during the experiment are presented in the next section of the study.

Results and discussion

As a result of the “methodology for studying the attitude of students in the 3rd grade to the subject of study” by G.N. Kazantseva, three main levels of motivation of students were identified:

High level. 10-15 points – high level of motivation and educational activity to study, learn mathematics. Such children will have cognitive abilities and the desire to successfully fulfill all the requirements set by the school. Students carefully complete all tasks in mathematics, understand the essence of tasks and the main place of mathematics in the life of everyone and themselves.

Medium level. 5-9 points – good motivation to study mathematics. Most of the primary school children have such indicators and are successfully engaged in educational activities. The students tried to carefully complete the math assignments, understand the essence of the assignments and the main place of mathematics in the life of each person and themselves.

Low level. 1-4 points – primary school students have a low motivation to learn or study mathematics. Such children are poorly motivated in mathematics lessons, they want the lesson to be canceled. There are significant difficulties in educational activities.

G.N. Kazantseva's “methodology for studying attitude to the subject of study” made it possible to determine the level of motivation of Primary School students (Table 4).

Table 4 – Percentage indicators of the motivation level of primary school children “methodology for studying attitudes to an academic subject” by G.N. Kazantseva

groups	Indicators					
	Low	n	Medium	n	high	n
Control n=31	29%	9	45,2 %	14	25,8%	8
Experimental n=32	31,2%	10	43,8%	14	25 %	8

According to the results of the initial stage (identification): in the control group, the proportion of low-motivated students was 29% (9 students), while in the experimental group this figure increased slightly to 31.2% (10 students). This indicates that there is a significant proportion of students with low motivation for mathematics in both groups. The proportion of moderately motivated students in both groups is 14 students, which are 45.2% and 43.8%. The proportion of highly motivated students in both groups is similar: 8 students, which is 25.8% in the control group, 25% in the experimental group. This indicates that before the start of the study, the number of students with high levels of motivation for mathematics in both groups was approximately equal.

Initial results obtained according to G.N. Kazantseva's methodology shows that elementary school students have different levels of motivation for mathematics. Although the average level of motivation prevails, the proportion of low-motivated students is also significant. The proportion of highly motivated students is relatively small in both groups. This indicates the importance of work aimed at increasing motivation in the formative stage of study.

The results of the control and experimental groups are presented in the form of a diagram below (Figure 1).

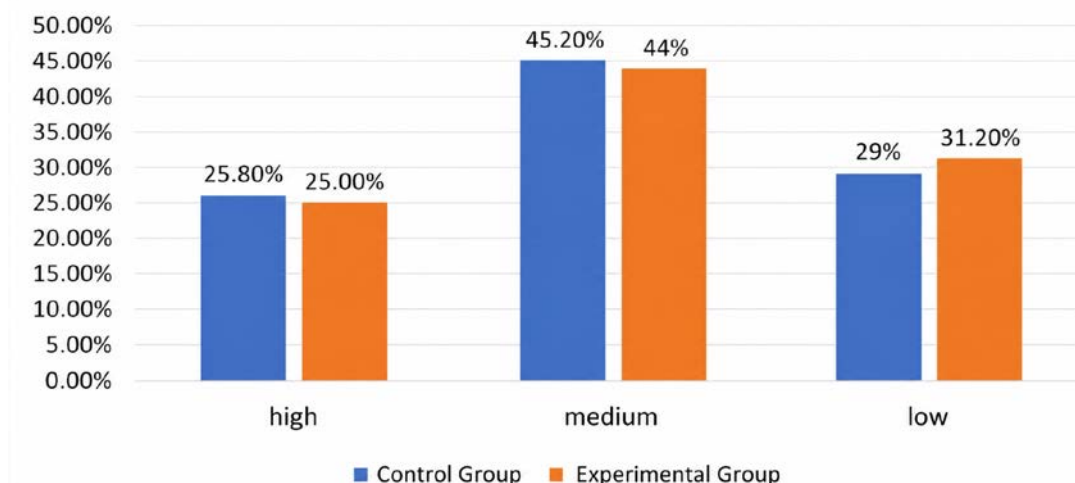


Figure 1 – Histogram of the Percentage Indicators of Primary School Students' Motivation Levels Based on G.N. Kazantseva “Methodology for Studying Attitudes Toward a School Subject”

The questionnaire For determining academic activity and motivation in mathematics (author's survey) allowed us to obtain the following results:

1. *Do you like it when you publish a report?* 52% in the control groups and 53% in the experimental groups showed positive answers to the first question. Negative responses were 48% in the control groups and 47% in the experimental groups. Consequently, the enthusiasm of both groups for writing the report was half of the class.

2. *Do you want to solve problems in math class with interest and solve additional problems besides the textbook?* We see that the majority of both groups answered negatively, that is, those who answered “No” in the control groups showed 55%, and in the experimental groups-54%. Those who answered “Yes” showed 45% in the control groups and 46% in the experimental groups. From these indicators, it follows that, first of all, it is necessary to interest students in mathematics.

3. *Do you have any difficulties doing your math homework?* 41% in the control groups and 38% in the experimental groups showed positive answers to this question. Negative responses were 59% in the control groups and 62% in the experimental groups. We can see that the overwhelming majority answered this question in the negative. Therefore, first of all, we believe that if we explain the tasks to students when doing classwork and interest them in various methods, then we will not have problems with homework.

4. *Do you like to solve problems and complete tasks in a math textbook?* 48% in the control group and 51% in the experimental group showed positive answers to the question. Negative responses were 52% in the control groups and 49% in the experimental group. This shows that half of the respondents can perform research tasks in a textbook.

5. *Do you like writing a paper on a math research project and doing research work?* When asked “no”, the overwhelming majority chose the answer “no”, indicating 66% in the control group and 69% in the experimental group. We conclude that in order to interest students in project work in mathematics and to involve them in research work, it is necessary, first of all, to perform educational and cognitive research work performed in the classroom, solve additional logical research tasks, and develop research skills. And those who chose the answer “yes” showed 34% in the control groups and 31% in the experimental group.

6. *Would you like to work on interesting logical inquiry tasks that contribute to the development of research skills during mathematics lessons?* 61% of the control group and 58% of the experimental group answered this question affirmatively. This indicates a demand among students for interesting logical inquiry tasks in mathematics.

The table below shows the results of the survey conducted during the ascertaining experiment (Table 5).

Table 5 – Survey questions and results

Questions	Answers			
	Control group		Experimental group	
	yes	no	yes	no
1. Do you like Mathematics?	52%	48%	47%	53%
2. Do you want to solve problems in mathematics class with interest and solve additional problems in addition to the textbook?	45%	55%	46%	54%
3. Are you having trouble doing your math homework?	41%	59%	38%	62%
4. Do you like to solve problems on the conditional sign in the textbook and perform tasks in Mathematics?	48%	52%	51%	49%
5. Do you like to write a research project in Mathematics and engage in research work?	34%	66%	31%	69%
6. Do you want mathematics to have interesting logical-research problems and tasks that form the skill of research?	61%	39%	58%	42%

The results of the survey of control and experimental groups are shown in Figures 2, 3 in the form of a diagram.

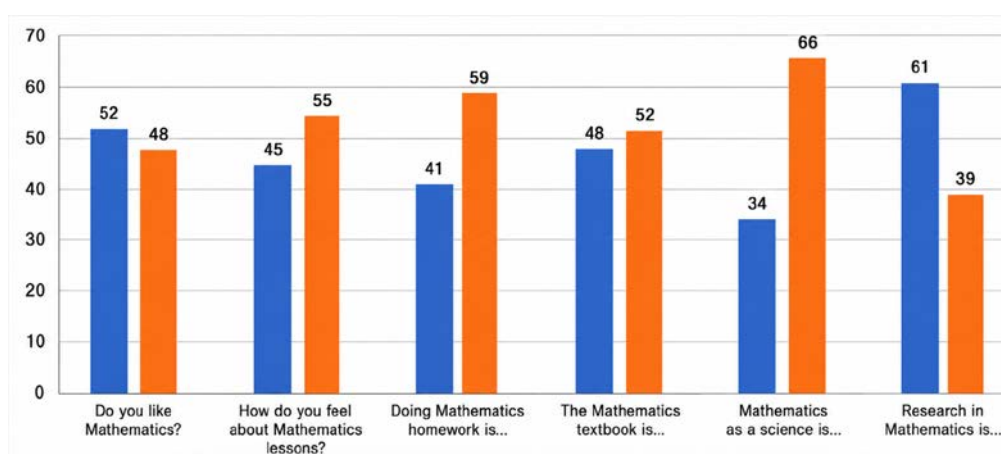


Figure 2 – Control group survey results

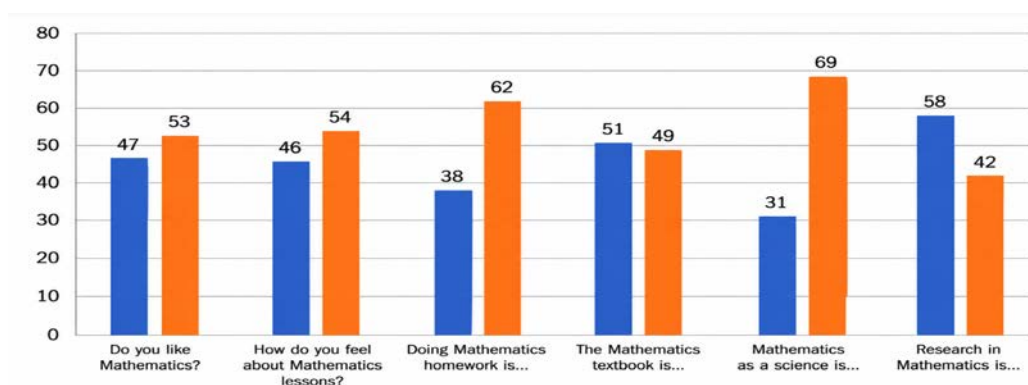


Figure 3 – Experimental group survey results

After analyzing the survey results, the following conclusions can be drawn:

General interest in mathematics (Question 1):

The proportion of mathematics enthusiasts in the control group (52%) is higher than in the experimental group (47%). This shows that the overall positive attitude towards mathematics slightly prevails in the control group.

Enthusiasm for solving additional tasks (Question 2): in both groups, the proportion of those who do not have the motivation to solve additional tasks prevails (55% and 54%). This suggests that motivation to solve additional extracurricular tasks is not high in both groups.

Difficulties with homework (Question 3): In both groups, the proportion of those who have difficulty completing homework is high (59% and 62%). This indicates that students have difficulty completing their math homework. In the experimental group, this indicator is slightly higher.

Sympathy for solving problems in the textbook (question 4): The answers to this question in the two groups are very similar. The sympathy for solving problems in the textbook is approximately equal in both groups.

Interest in research (Question 5): Both groups have low interest in research (34% and 31%). This suggests that younger students generally have low motivation to engage in scientific projects.

The need for tasks that develop research skills (question 6): On this issue, the proportion of those in both groups who answered “Yes” to the need for tasks that develop research skills is higher (61% and 58%). These findings suggest that students are inclined to incorporate research elements into mathematics lessons. The analysis of the survey results obtained during the ascertaining experiment led us to conclude that specific measures are required to enhance primary school students’ learning activity and motivation to study mathematics.

According to the content component of the ascertaining experiment, A. I. Savenkov made it possible to determine the quantitative and qualitative indicators of the levels of students' research skills using the methodology “determining students' research skills” (Table 6).

Table 6 – Percentage indicators of students' research skills derived from the methodology of A.I. Savenkov “determining students' research skills” on the content component of the detection experiment

Groups	Low	Indicators				
		n	Medium	n	High	n
Control n=31	38,7 %	12	48,3%	15	13%	4
Experimental n=32	40,6%	13	43,8%	14	15,6%	5

The results of the ascertaining experiment conducted in accordance with the present methodology showed that:

Low level: in the control group, the low level of research skills was 38.7% (12 students), in the experimental group this indicator was increased to 40.6% (13 students). This may mean that in the experimental group, the initial level of research skills decreased slightly compared to the control group or she had problems with its formation.

Average level: in the control group, the average level of students was 48.3% (15 students), in the experimental group this figure decreased to 43.8% (14 students). This also indicates a decrease in the proportion of middle-level students in the experimental group.

High level: the proportion of students demonstrating a high level of research skills was 13% (4 students) in the control group, compared with 15.6% (5 students) in the experimental group. The difference indicates a modestly higher proportion of students reaching the high level in the experimental group.

The results of the control and experimental groups are presented below in the form of a diagram (Figure 4).

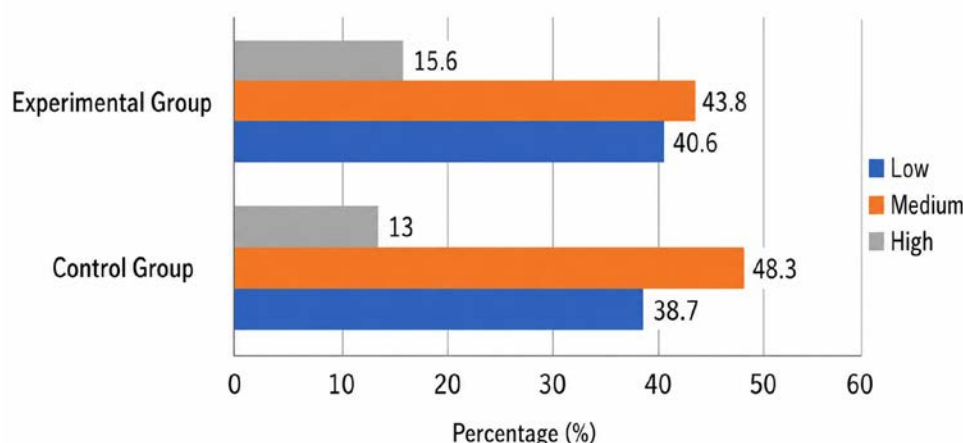


Figure 4 – The results of diagnostics according to the methodology of A.I. Savenkov determining the research skills of primary school children

It can be seen from the picture that most of the students are intermediate and low level. Low motivation of students in the field of research the figure shows that the majority of students are at an average and low level, as well as the fact that they act only under the guidance of a teacher, are not ready for independent activity-indicates the need for students to develop research skills.

Summing up the results of the stage of the study of the substantive component of the experiment by A.I. According to Savenkov, the results of the methodology “determining students' research skills” show that in the experimental group, some indicators of the initial level of research skills (at low and medium levels) slightly deteriorated compared with the control group, and the proportion of students at a higher level increased slightly.

Receiving the first experimental data, the first experimental data on the results were obtained at the initial stage of this experimental work, later in the formative experiment, a set of skills is organized that allow students of the experimental group to independently search for new knowledge, as well as solving problems in teaching mathematics to elementary grades and activities, In order to increase the dominance of internal cognitive motives of students' educational and research work, it is necessary to develop a methodological system for the formation of students' research skills.

According to Table 7, the following results of the survey “my attitude to research activities” were obtained.

Table 7 - The obtained percentage indicators of the survey by Y.A. Kazimirova “my attitude to research activities”

Groups	Indicators					
	Low	n	medium	n	high	n
Control n=31	32,3%	10	51,6%	16	16,1%	5
Experimental n=32	40,6%	13	46,9%	15	12,5%	4

According to the results of the modified survey by Yu.A. Kazimirova, there is a decrease in the level of general interest in research activities in the experimental group. In particular, the proportion of low-level students has increased, while the proportion of middle and high-level students has decreased. The study shows that more than a third of students lack the knowledge, understanding and specific skills of research activities. In the experimental group as a whole, there may be a decrease in interest in research activities compared to the control group or the presence of certain difficulties in its development. In this regard, it can be shown that the research work carried out did not have a positive effect on increasing students' interest in research in the experimental group and

even had the opposite effect. Therefore, the results may indicate the need to revise methods and approaches in future work.

The results of the survey “my attitude to research” (Yu.A. Kazimirova) are shown in Figure 5.

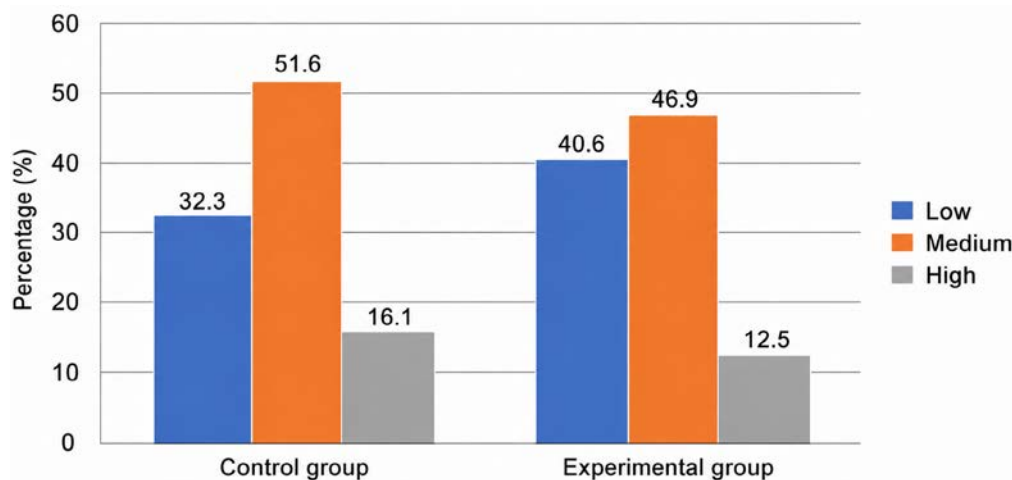


Figure 5 – Diagnostic results according to the method of Yu.A. Kazimirova “my attitude to research activities” at the stage of defining the experiment

The results obtained during the research method, selected according to the reflexive component, are presented in Table 8.

Table 8 - Results of the methodology for determining the reflexive level of students according to the “methodology for diagnosing the level of development of reflexivity” by A.V. Karpov

Levels	Groups			
	Control group n=31		Experimental group n=32	
	%	n	%	n
Low level of reflexivity development	22,6 %	7	21,8%	7
The average level of development of reflexivity	29 %	9	31,3%	10
High level of reflexivity development	48,4%	15	46,9%	15

Table 8 presents the results obtained using the “methodology for diagnosing the level of reflection development”. It shows that the proportion of students demonstrating a low level of reflection is roughly the same in both groups (22.6% in the control group and 21.8% in the experimental group). This may indicate that these students find it difficult to critically evaluate their own actions.

The percentage of students in the experimental group demonstrating a medium level of reflection (31.3%) is slightly higher than that of the control group (29%). This means that reflexivity is higher in the experimental group than in the control group.

In both groups, the proportion of students with high reflexivity is approximately the same (48.4% in the control group and 46.9% in the experimental group). This suggests that the experimental work did not significantly increase the high level of reflexivity of the students.

The results obtained using A.V. Karpov's methodology indicate the predominance of high-level reflexivity in younger schoolchildren, especially in the process of developing research skills in

mathematics (more than 45% in both groups). This means that students have a high ability to analyze their actions. The low level of reflection is relatively low in both groups.

Although there was a slight increase in the proportion of students with average reflexivity in the experimental group, there was no significant difference between the groups in terms of the overall level of high-level reflexivity. This may indicate that the actions aimed at developing reflection are still small or that they do not have a rapid effect on the already formed level of reflexivity of students.

Thus, the first stage of defining the experimental work was done as follows:

- Firstly, in the control and experimental groups on the “methodology for studying attitudes to an academic subject” in order to determine the level of motivation of younger schoolchildren for an academic subject, there is not much difference in all levels of motivation of students to study and study mathematics.

- Secondly, the “Questionnaire for identifying academic activity and motivation in mathematics” based on the results of the author's survey of primary school students in order to identify academic activity and motivation in mathematics showed low interest in writing research papers in mathematics in control and experimental groups.

- Thirdly, as the diagnostic card “identification of students' research skills” shows, the low motivation of students in the field of research, as well as the fact that they act only under the guidance of a teacher, are not ready for independent activity, indicates the need for students to develop research skills.

- Fourthly, as the results of the survey “my attitude to research” show, it is clear that students do not attempt research, are inactive, often neglect it, and more than a third of students do not have any knowledge, understanding and specific skills of research activity. This indicates the need for students to develop research skills.

Fifth, according to the results of the reflexive level, both groups had almost the same levels of students, and there were no very high-level students in both groups who could reflect on their actions.

Conclusion

Pedagogical diagnostics allowed us to collect reliable information about the state of formation of research skills of 3rd grade students in mathematics lessons.

Despite the positive results achieved, we cannot say that the research problem has been fully resolved. This is because developing research skills in primary school students is a complex process requiring long-term, systematic effort. Therefore, additional comprehensive research is needed to study the most effective pedagogical methods and conditions for the development of research skills in younger schools.

Based on the results of the study, the following recommendations were formulated:

- systematic inclusion of problematic situations, logical tasks, and research tasks in the content of primary school mathematics lessons in order to develop students' research skills;

- development of a system of teaching aids and assignments for organizing student research for primary school teachers.

It cannot be said that the research problem was completely solved, because it was very difficult. It is necessary to continue a comprehensive study of the problem of forming the research skills of elementary school children with formative research in order to implement not only in mathematics lessons, but also in any primary school subject and improve the results obtained. Future research may aim to expand the implementation of training research activities not only in math lessons, but also in other primary school subjects. In addition, it is important to analyze the long-term effects of student-based learning academic performance, critical thinking, problem solving skills, and lifelong learning skills. Research contributes to the improvement of educational practice and the improvement of the overall quality of primary education.

Expression of gratitude

We would like to express our gratitude to the management and staff of the E. Izatullaev Comprehensive School in Turkestan, Turkestan region, who jointly conducted an empirical study.

References

1. Қазақстан Республикасының «Білім туралы» Заңы. 27 шілде 2007 ж., №319-III ҚРЗ (2025.11.01. берілген өзгертулермен және толықтырулармен) <https://adilet.zan.kz/kaz/docs/Z070000319>
2. Мектепке дейінгі тәрбие мен оқытудың, бастауыш, негізгі орта, жалпы орта, техникалық және кәсіптік, орта білімнен кейінгі білім берудің мемлекеттік жалпыға міндетті стандарттары. Қазақстан Республикасы Оқу-ағарту министрінің 2025 жылғы 23 қаңтардағы, №12 бұйрығы <https://adilet.zan.kz/kaz/docs/V2200029031>
3. Жалпы білім беру ұйымдарына арналған жалпы білім беретін пәндердің, бастауыш, негізгі орта және жалпы орта білім деңгейлерінің таңдау курстарының үлгілік оқу бағдарламалары. Қазақстан Республикасы Оқу-ағарту министрінің 2023 жылғы 31 қазандағы, №328 бұйрығы <https://adilet.zan.kz/kaz/docs/V2200029767>
4. Қазақстан Республикасында мектепке дейінгі, орта, техникалық және кәсіптік білім беруді дамытудың 2023 - 2029 жылдарға арналған тұжырымдамасы. Қазақстан Республикасы Үкіметінің 2023 жылғы 28 наурыздағы № 249 қаулысы <https://adilet.zan.kz/kaz/docs/P2300000249>
5. «Орталық Азия-Қытай» екінші саммиті https://t.me/aqorda_resmi
6. Ризаева Л.А. Бастауыш сынып математикасын оқытуда есептерді шығару арқылы оқушылардың зерттеу дағдыларын қалыптастыру: филос. д-ры. дисс. - Шымкент, 2025. - 186 б.
7. Rizayeva L., Kenzhebekova R., Amirova A., Akhatayeva U., Uaidullakzy E., Smanova B. Formation of research skills of students through solving problems in teaching mathematics in primary classes. //Cypriot Journal of Educational Sciences, 2022. - 17(8). <https://doi.org/10.18844/cjes.v17i8.7824>
8. Fry K., Nakar S., Zorn K. Professional learning interventions for inquiry-based pedagogies in primary classrooms: A scoping review. // Mathematics Education Research Journal, 2025. - 37. - 853-887 <https://doi.org/10.1007/s13394-024-00516-x>
9. Нишанбаева С.Х., Сәдуақас Г.Т., Абилхаирова Ж.А. Оқыту үдерісіндегі ғылыми түсініктер арқылы бастауыш сынып оқушыларының зерттеушілік мәдениетін қалыптастыру. //Абай атындағы ҚазҰПУ-ң Хабаршысы «Педагогика ғылымдары» сериясы. - 2021. - № 4 (72). <https://doi.org/10.51889/2021-4.1728-5496.33>
10. Kurebay B., Mailybayeva G., Sharipkhojayeva Z., Zharkumbayeva M., Kozhamkulova N. Digital technology integration for developing research competence in primary school students. // Front. Educ. 2026. - 11:1821816. <https://doi.org/10.3389/educ.2026.1821816>
11. Bapanova G.K., Orekhova N.V., Kadirsizova S.B., Kasbayeva G.S., Sholpankulova G.K. Research Skills as the Student Learning Achievement. //International Journal of Educational Reform. - 2025. - 34(3). - 543-558. <https://doi.org/10.1177/10567879231155874>
12. López Ch. Educational model to strengthen skillsresearch in primary school students. // University, Science and Technology. Special Issue. - 2024. - pp. 152-162 <https://doi.org/10.47460/uct.v28iSpecial.785>
13. Sky Ariella. The Most Important Research Skills (With Examples). - 2023 <https://www.zippia.com/advice/research-skills/>
14. Исаева З.А. Формирование профессионально-исследовательской культуры педагога в системе университетского образования: дис. док. пед. наук: 13.00.01. - Алматы, 1997. - 334 с.
15. Туралбаева К.Т., Джарасова Г.С., Жуматова И.Ж., Сагнаева Г.Т., Асимова Б.К. Математика пәні бойынша тесттің психометриялық талдауы: раш моделі негізінде бағалау. //Педагогикалық өлшемдер. 2026. - №1(5). - Б. 33-41 <https://doi.org/10.63597/UTO3105-4161.2026.5.1.003>

Л. Ризаева, А. Жунисова, Л. Нурсеитова, К.Т. Туралбаева, Л. Сейсекенова

МАТЕМАТИКА САБАҚТАРЫНДА БАСТАУЫШ СЫНЫП ОҚУШЫЛАРЫНЫҢ ЗЕРТТЕУ DAҒДЫЛАРЫН ҚАЛЫПТАСТЫРУ ДИАГНОСТИКАСЫ

Мақалада математика сабақтарында бастауыш сынып оқушыларының зерттеу дағдыларын қалыптастыру мәселесі қарастырылған. Нормативтік құжаттарға сүйене отырып, зерттеу өзектілігі айқындалған. Зерттеу барысында теориялық әдіс арқылы отандық және шетелдік зерттеулер бойынша зерттеу дағдысы ұғымының мәні айқындалып, зерттеу дағдысының артықшылықтары сипатталған. Диагностикалық әдістер бастауыш сынып оқушыларының зерттеу дағдыларын қалыптастырудың мотивациялық, мазмұндық және рефлексивтік компоненттеріне сәйкес іріктелген. Эмпирикалық зерттеу нәтижелері әр компонент бойынша оқушылардың зерттеу дағдыларының даму деңгейін жақ-жақты диагностикалауға мүмкіндік берген. Алынған сандық деректерге сүйене отырып, бастауыш сынып оқушыларының зерттеу дағдыларының төмендегіне байланысты қалыптастырушы зерттеу кезеңінің қажет екендігін тұжырымдаған. Зерттеу қорытындысында ғылыми негізделген келесідей ұсыныстар берілген: бастауыш сыныптың математика сабақтарының мазмұнына оқушылардың зерттеу дағдыларын дамыту мақсатында проблемалық жағдаяттар, логикалық тапсырмалар, зерттеушілік сипаттағы тапсырмаларды жүйелі енгізу; бастауыш сынып мұғалімдеріне арналған оқушылардың зерттеушілік жұмыстарын ұйымдастыру бойынша әдістемелік құралдар мен тапсырмалар жүйесін әзірлеу. Математика сабақтарында бастауыш сынып оқушыларының зерттеу дағдыларын қалыптастырудың диагностикалануы мен ғылыми зерттеу нәтижелерін жоғары педагогикалық оқу орындары мен колледждерде, бастауыш сынып мұғалімдерін даярлау және қайта даярлау біліктілікті арттыру институттарында пайдалануға болатыны бұл зерттеу жұмысының практикалық маңыздылығын көрсетеді.

Түйін сөздер: зерттеу дағдысы, математика, қызығушылық, жобалау, логикалық ойлау, шығармашылық, диагностика.

Л. Ризаева, А. Жунисова, Л. Нурсеитова, К.Т. Туралбаева, Л. Сейсекенова

ДИАГНОСТИКА РАЗВИТИЯ ИССЛЕДОВАТЕЛЬСКИХ УМЕНИЙ У МЛАДШИХ ШКОЛЬНИКОВ НА УРОКАХ МАТЕМАТИКИ

В статье рассматривается проблема формирования исследовательских умений у младших школьников на уроках математики. Актуальность исследования обоснована с опорой на нормативные документы. В рамках теоретического подхода раскрывается сущность понятия «исследовательские умения» на основе анализа отечественных и зарубежных исследований, а также описывается значимость данных умений. Методы диагностики были подобраны в соответствии с мотивационным, содержательным и рефлексивным компонентами процесса формирования исследовательских умений у учащихся начальной школы. Результаты эмпирического исследования позволили всесторонне оценить уровень развития исследовательских умений учащихся по каждому из компонентов. На основе полученных количественных данных был сделан вывод о необходимости проведения формирующего этапа работы, направленного на развитие исследовательских умений у учащихся начальной школы. На основе результатов исследования предлагаются следующие научно обоснованные рекомендации: систематическое включение в уроки математики в начальной школе проблемных ситуаций, логических задач и заданий исследовательского характера для формирования у учащихся исследовательских умений, а также разработка методических материалов и системы заданий для учителей начальных классов, способствующих организации исследовательской деятельности учащихся. Практическая значимость данного исследования подтверждается тем, что методы оценки развития исследовательских умений у младших школьников на уроках математики – наряду с самими результатами исследования – могут быть использованы в педагогических вузах и колледжах, а также в учреждениях, занимающихся подготовкой, переподготовкой и повышением квалификации учителей начальных классов.

Ключевые слова: исследовательские навыки, математика, интерес, проектная деятельность, логическое мышление, креативность, диагностика.

References

1. Parliament of the Republic of Kazakhstan. (2007). *Qazaqstan Respublikasynyn «Bilim turaly» Zany* [Law of the Republic of Kazakhstan «On Education»] (Law No. 319-III, as amended on November 1, 2025). Adilet Information and Legal System of Regulatory Legal Acts of the Republic of Kazakhstan. <https://adilet.zan.kz/kaz/docs/Z070000319>
2. Ministry of Education of the Republic of Kazakhstan. (2025). *Mektepke deingi tarbie men oqytudyn, bastauysh, negizgi orta, zhalpy orta, tekhnikalyq zhane kasiptik, orta bilimnen keingi bilim berudyn memlekettik zhalpyga mindetti standarttaryn bekitu turaly* [On approval of state compulsory standards of preschool education and training, primary, basic secondary, general secondary, technical and vocational, post-secondary education] (Order No. 12). Adilet Information and Legal System of Regulatory Legal Acts of the Republic of Kazakhstan. <https://adilet.zan.kz/kaz/docs/V2200029031>
3. Ministry of Education of the Republic of Kazakhstan. (2023). *Zhalpy bilim beru uyymdaryna arналған zhalpy bilim beretin padinderdyn, bastauysh, negizgi orta zhane zhalpy orta bilim dengeilerynyn tandalau kurstarynyn ulgilik oqu bagdarlamalaryn bekitu turaly* [On approval of model curricula for general education subjects and elective courses for primary, basic secondary, and general secondary education levels] (Order No. 328). Adilet Information and Legal System of Regulatory Legal Acts of the Republic of Kazakhstan. <https://adilet.zan.kz/kaz/docs/V2200029767>
4. Government of the Republic of Kazakhstan. (2023). *Qazaqstan Respublikasynda mektepke deingi, orta, tekhnikalyq zhane kasiptik bilim berudi damytudyn 2023–2029 zhyldarga arналған tuzhyrymdamasy* [Concept for the development of preschool, secondary, technical, and vocational education in the Republic of Kazakhstan for 2023–2029] (Decree No. 249). Adilet Information and Legal System of Regulatory Legal Acts of the Republic of Kazakhstan. <https://adilet.zan.kz/kaz/docs/P2300000249>
5. Press Service of the President of the Republic of Kazakhstan (Akorda). (n.d.). «Ortalyq Azia–Qytai» ekinshi sammiti [The Second «Central Asia–China» Summit] [Telegram channel]. Telegram. Retrieved September 28, 2026, from https://t.me/aqorda_resmi
6. Rizayeva, L. A. (2025). *Bastauysh synyp matematikasy oqytuda esepтерdi shygaru arqyly oqushylardyn zertteu dagdylaryn qalyptastyru* [Formation of research skills of students through solving problems in teaching primary school mathematics] (Doctoral dissertation / PhD thesis). Shymkent
7. Rizayeva, L., Kenzhebekova, R., Amirova, A., Akhatayeva, U., Uaidullakzy, E., Smanova, B. (2022). Formation of research skills of students through solving problems in teaching mathematics in primary classes. *Cypriot Journal of Educational Sciences*, 17(8). <https://doi.org/10.18844/cjes.v17i8.7824>
8. Fry, K., Nakar, S., Zorn, K. (2025). Professional learning interventions for inquiry-based pedagogies in primary classrooms: A scoping review. *Mathematics Education Research Journal*, 37, 853–887 <https://doi.org/10.1007/s13394-024-00516-x>
9. Nishanbayeva, S. Kh., Saduakas, G. T., & Abilkhairova, Zh. A. (2021). *Oqytu uderisindei ghylymi tusinikter arqyly bastauysh synyp oqushylardyn zertteushilik madenietin qalyptastyru* [Formation of research culture of primary school students through scientific concepts in the learning process]. *Abai Kazakh National Pedagogical University Bulletin, «Pedagogical Sciences» Series*, 4(72), 1728–5496. <https://doi.org/10.51889/2021-4.1728-5496.33>
10. Kurebay, B., Mailybayeva, G., Sharipkhojayeva, Z., Zharkumbayeva, M., Kozhamkulova, N. (2026). Digital technology integration for developing research competence in primary school students. *Front. Educ*, 11:1821816. <https://doi.org/10.3389/educ.2026.1821816>
11. Bapanova, G.K., Orekhova, N.V., Kadirsizova, S.B., Kasbayeva, G.S., Sholpankulova, G.K. (2025). Research Skills as the Student Learning Achievement. *International Journal of Educational Reform*, 34(3), 543–558. <https://doi.org/10.1177/10567879231155874>
12. López, Ch. (2024). Educational model to strengthen skillsresearch in primary school students. *University, Science and Technology. Special Issue*, 152–162 <https://doi.org/10.47460/uct.v28iSpecial.785>
13. Sky, Ariella (2023). The Most Important Research Skills (With Examples). <https://www.zippia.com/advice/research-skills/>
14. Issayeva, Z. A. (1997). *Formirovanie professional'no-issledovatel'skoj kul'tury pedagoga v sisteme universitetskogo obrazovaniya* [Formation of professional and research culture of a teacher in the system of university education] (Doctoral dissertation, Doctor of Pedagogical Sciences). Almaty
15. Turalbayeva, K. T., Djarasova, G. S., Zhumatova, I. Zh., Sagnayeva, G. T., & Asemova, B. K. (2026). *Matematika pani boinsha testtin psikhometrialyq taldaу: Rash modeli negizinde bagalau*

[Psychometric analysis of a mathematics test: Evaluation based on the Rasch model]. Pedagogical Measurements, 1(5), 33–41. <https://doi.org/10.63597/UTO3105-4161.2026.5.1.003>

Авторлар туралы мәлімет:

Ризаева Луиза – аға оқытушы, PhD (философия докторы) Ж.А. Тәшенев атындағы университет, Шымкент, Қазақстан, e-mail: luizarizaeva84@gmail.com

Жунисова Акнур (автор-корреспондент) - аға оқытушы, PhD (философия докторы) Ж.А. Тәшенев атындағы университет, Шымкент, Қазақстан, e-mail: aknura_er74@mail.ru

Нурситова Лаззат – аға оқытушы, докторант, Ж.А. Тәшенев атындағы университет, Шымкент, Қазақстан, e-mail: dala1971@mail.ru, 87028483205

Туралбаева Кенжекул Турганбаевна – зертхана меңгерушісі, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail: kengebahit@mail.ru

Сейсекенова Лаура – магистрант, Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан, e-mail: seisekenova.laura.1973@mail.ru

Information about authors:

Rizayeva Luiza – Senior Lecturer, PhD (Doctor of Philosophy), Zh.A. Tashenev University, Shymkent, Kazakhstan, e-mail: luizarizaeva84@gmail.com

Zhunissova Aknur (corresponding author) – Senior Lecturer, PhD (Doctor of Philosophy), Zh.A. Tashenev University, Shymkent, Kazakhstan, e-mail: aknura_er74@mail.ru

Nursitova Lazzat - Senior Lecturer and Doctoral Student, Zh.A. Tashenev University, Shymkent, Kazakhstan, Shymkent, Kazakhstan, e-mail: dala1971@mail.ru

Turalbayeva Kenzhukul Turganbayevna – Head of laboratory, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan, e-mail: kengebahit@mail.ru

Seisekenova Laura – master's student, Kazakh National Pedagogical University named after Abay, Almaty, Kazakhstan, e-mail: seisekenova.laura.1973@mail.ru

Сведения об авторах:

Ризаева Луиза – старший преподаватель, PhD (доктор философии), Университет имени Ж.А. Ташенева, Шымкент, Казахстан. e-mail: luizarizaeva84@gmail.com

Жунисова Акнур (автор-корреспондент) – старший преподаватель, PhD (доктор философии), Университет имени Ж.А. Ташенева, Шымкент, Казахстан, e-mail: aknura_er74@mail.ru

Нурситова Лаззат – старший преподаватель, докторант, Университет имени Ж.А. Ташенева, Шымкент, Казахстан, e-mail: dala1971@mail.ru

Туралбаева Кенжекул Турганбаевна – заведующий лабораторией, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail: kengebahit@mail.ru

Сейсекенова Лаура – магистрант, Казахский Национальный педагогический университет имени Абая, Алматы, Казахстан, e-mail: seisekenova.laura.1973@mail.ru

Р.С. Қалиева¹, Г.Ш.Ерназарова², М.Н.Қойшыбаев³, К.С.Кенжебаева^{4*}

^{1,2,3,4}Коркыт Ата атындағы Қызылорда университеті, Қызылорда қ., Қазақстан

*e-mail: kundiz_27_87@mail.ru

¹ORCID 0009-0008-4675-7849, ²ORCID 0000-0002-4516-3415

³ORCID 0000-0002-4650-8630, ⁴ORCID 0000-0003-0908-6843

БАСТАУЫШ СЫНЫП ОҚУШЫЛАРЫНЫҢ КІТАП ОҚУҒА ҚЫЗЫҒУШЫЛЫҒЫН АРТТЫРУДА ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАРДЫ ҚОЛДАНУ: МҰҒАЛІМДЕРДІҢ ТӘЖІРИБЕСІ МЕН БАҒАЛАУЫ

Бастауыш сынып оқушыларының кітапқа қызығушылығы төмендеген кезде цифрлық құралдың сыныпта болуы мәселені өздігінен шешпейді. Интерактивті тақта орнатылып, электрондық оқулық қолжетімді болғанымен, мұғалім оларды оқу мотивациясын күшейту үшін қаншалықты жиі әрі қалай қолданады деген сұрақ ашық қалып отыр. Қызылорда облысында жүргізілген зерттеу осы алшақтықты нақтылауға бағытталды. Сандық сауалнамаға бастапқыда 156 бастауыш сынып мұғалімі қатысты. Деректерді тексеру кезінде талапқа сай келмеген 7 жауап алынып тасталып, талдауға 149 мұғалімнің мәліметі енгізілді. Сауалнама цифрлық технологияларды қолдану жиілігін ғана емес, мұғалімдердің олардың ықпалына берген бағасын, цифрлық даярлығын, техникалық мүмкіндіктерін және күнделікті сабақта кездесетін кедергілерді де қамтыды. Мәліметтер IBM SPSS Statistics 29 бағдарламасында өңделді. Айнымалылар арасындағы байланысты тексеру үшін Спирменнің рангтік корреляциясы, ал мұғалімдер топтарының айырмашылықтарын салыстыру үшін Краскел –Уоллис критерийі қолданылды. Нәтиже бір қайшылықты көрсетті: мұғалімдер цифрлық құралдардың педагогикалық мүмкіндігін жоғары бағалайды, бірақ оларды сабақта тұрақты қолдана бермейді. Ең жиі пайдаланылатын құралдар – электрондық оқулық пен интерактивті тақта. Алайда құрылғының болуы оның мультимодалдық мүмкіндіктері, соның ішінде мәтін, дыбыс, бейне және интерактивті тапсырмаларды бір оқу әрекетіне біріктіру толық пайдаланылып жүр дегенді білдірмейді. Мәселе техниканы көбейтумен шектелмейді. 149 мұғалімнің жауаптары цифрлық құралды оқу мотивациясымен байланыстыратын нақты әдістемелік қолдау қажет екенін көрсетеді. Сондықтан мұғалімдердің цифрлық даярлығын жетілдіру, сабаққа бейімделген нұсқаулықтар ұсыну және технологияны қолдануға кедергі келтіретін факторларды бөлек қарастыру отандық бастауыш білім беру тәжірибесі үшін маңызды міндет болып қала береді.

Түйін сөздер: бастауыш сынып, кітап оқуға қызығушылық, цифрлық технологиялар, оқу мотивациясы, мұғалімдердің даярлығы, электрондық оқулықтар, интерактивті оқыту.

Кіріспе

Бастауыш мектепте бала әріпті танып қана қоймайды, дәл осы кезеңде кітапты өз еркімен аша ма, әлде оқуды тек берілген тапсырма ретінде қабылдай ма – сол әдет қалыптасады[1]. Бірақ қызығушылық бір деңгейде сақталмайды. Әсіресе 4-сыныптан кейін оқуға деген ынта төмендей бастайтыны бірнеше зерттеуде тіркелген [2, 3]. Мұнда үш ұғымды араластырып алмау керек. Оқу мотивациясы баланы оқуға итермелейтін ішкі не сыртқы себепті және оның кітапқа қанша уақыт бөлетінін білдіреді [4]. Оқу қызығушылығы мәтінге немесе оқу әрекетіне қатысты эмоциялық ұнатумен байланысты [5]. Ал оқуға тартылу одан күрделірек, оқушы мәтінге зейін қойып, мағынаны алдын ала болжайды, оқиғаны көз алдына елестетеді, негізгі ойды жинақтайды және осы стратегияларды мұғалімнің тұрақты нұсқауынсыз қолдана бастайды [6]. Демек, бала кітапты ұнатуы мүмкін, бірақ мәтінмен терең жұмыс істемеуі де ықтимал. Оқу мотивациясын тек «оқуға қызығады» деген бір сұрақпен өлшеу сондықтан даулы.

Қазір бастауыш сыныптағы оқу мен жазуды интерактивті тақтасыз, электрондық кітапсыз немесе білім беру платформасынсыз елестету қиын. Дегенмен, құрылғының сыныпта тұруы мен оның оқу мақсатына қызмет етуі бір нәрсе емес. Цифрлық құрал мәтінді,

дыбысты және бейнені бір тапсырмаға біріктіре алады; жауапты бірден көрсетеді; тапсырманың күрделілігін оқушы деңгейіне қарай өзгертуге мүмкіндік береді. Ерте балалық шақтан 5-сыныпқа дейін жүргізілген зерттеулер мобильді және цифрлық технологиялардың фонетикалық түсінікті, сөздік қорды, мәтінді түсіну мен жазу дағдыларын қолдай алатынын көрсетеді [7]. Негізгі шарт – технология сабаққа жай қосымша емес, нақты оқу әрекетінің бөлігі болуы керек. Электрондық кітапта бір бетке жазбаша мәтін, диктор дауысы, иллюстрация және анимация қатар орналаса алады. Мұндай орта баланың назарын тез тартады. Бірақ назардың экранға ауғаны мәтіннің түсінілгенін білдірмейді. Метаталдау нәтижесінде мазмұнмен тікелей байланысқан анимация, музыка және дыбыстық әсерлердің мәтінді түсіну мен сөздік қорына шағын оң ықпал беретіні анықталған [8]. Ал оқиғаға қатысы жоқ ойын, жыпылықтаған батырма немесе артық интерактив баланың назарын негізгі желіден бөледі. Танымдық жүктеме өседі. Сондықтан интерактивті кітаптың әсерін «цифрлық» деген атауы емес, экрандағы әр элементтің не үшін берілгені анықтайды [9]. Экраннан оқу қағаздан оқуды автоматты түрде басып озған жоқ. Ақпараттық және құрылымы күрделі мәтіндерде қағаз форматтың шағын артықшылығы сақталады. Цифрлық және баспа кітаптарын салыстырған 39 эксперименттік зерттеудің метаталдауына 1 812 бала қатысқан; нәтиже кітап форматына ғана емес, цифрлық дизайнға, ересектің көмегіне және қойылған оқу мақсатына тәуелді болған [1]. Бір құрылғыны күн сайын пайдалану да өздігінен сапа көрсеткіші емес. Егер оқушы экранда тек жауапты белгілеу немесе сөзді сәйкестендіру сияқты үстірт жаттығуларды орындаса, мәтінді түсіну нәтижесі төмендеуі мүмкін [10]. Ал ақпарат іздеу, деректі салыстыру немесе шағын зерттеу жобасын орындау үшін қолданылған технология басқа нәтиже береді. Шешуші таңдау мұғалімге тиесілі: қандай құрал, сабақтың қай кезеңінде және қандай мәтінмен қолданылатыны маңызды. Мұнда цифрлық даярлықпен қатар педагогикалық білім, техникалық қолжетімділік және технологияға көзқарас қатар әсер етеді. Алайда мұғалімнің цифрлық құралды қолдауы оны сабақта тұрақты әрі мақсатты қолданатынын білдірмейді. Швецияда 289 бастауыш сынып мұғалімі қатысқан сауалнамада респонденттердің 82%-ы оқу мен жазуды үйретуде цифрлық технологияны пайдалануға қызығатынын айтқан. Соған қарамастан, нақты оқу дағдыларын дамытуға арналған құралдарды қолдану жиілігі мұғалімдер арасында біркелкі болмаған [11]. Кедергі көбіне бір ғана себептен шықпайды. Мұғалімнің уақыты шектеулі болуы мүмкін; оқу бағдарламасына сай қазақ тіліндегі сапалы қосымша табылмауы ықтимал; мектептегі құрылғы саны сыныптағы оқушыға жетпеуі немесе интернет сабақ үстінде үзіліп қалуы мүмкін. Бұған цифрлық құзыреттіліктің әркелкілігі, нақты әдістемелік нұсқаулықтың аздығы және бірнеше платформаның арасынан оқу мақсатына сай құралды таңдау қиындығы қосылады. Сол себепті зерттеуде мұғалімдердің цифрлық даярлығы, техникалық мүмкіндігі және практикада кездесетін кедергілері бөлек көрсеткіштер ретінде қарастырылады. Халықаралық еңбектердің едәуір бөлігі электрондық кітап пен цифрлық құралдың сөздік қорға, оқу дағдысына және мәтінді түсінуге ықпалын эксперимент немесе метаталдау арқылы бағалайды. Мұндай зерттеулер әсердің бар-жоғын көрсетуге көмектеседі. Бірақ күнделікті сыныптағы жағдайды толық ашпайды: мұғалім қай құралды жиі таңдайды, оның оқушы қызығушылығына ықпалын қалай бағалайды және неге кейбір құралды мүлде қолданбайды? Қазақстандық бастауыш мектептер бойынша осы сұрақтарды өзара байланыстырып көрсететін эмпирикалық дерек аз. Әсіресе мұғалімнің оң көзқарасы, цифрлық даярлығы және мектептің техникалық жағдайы нақты қолдану жиілігіне қалай ұласатыны әлі де тексеруді қажет етеді. Зерттеу осы алшақтыққа назар аударады. Негізгі мақсат – бастауыш сынып мұғалімдерінің оқушыларды кітап оқуға тарту үшін цифрлық технологияларды қаншалықты жиі қолданатынын, олардың ықпалын қалай бағалайтынын, қандай техникалық мүмкіндікке ие екенін және сабақ барысында қандай кедергілермен бетпе-бет келетінін анықтау.

Талдау үш бағытта жүргізіледі. Алдымен цифрлық технологияларды қолдану жиілігі, мұғалімдердің даярлығы және техникалық мүмкіндігі өлшенеді. Кейін осы көрсеткіштердің өзара байланысы мен педагогикалық өтілге қарай айырмашылықтары тексеріледі. Соңында

мұғалімнің технологияны таңдауы мен тұрақты қолдануына бөгет болатын себептер қарастырылады. Зерттеу нысаны – бастауыш мектеп оқушыларының кітап оқуға қызығушылығын арттыру үдерісі. Зерттеу пәні – мұғалімдердің осы мақсатта цифрлық технологияларды қолдану тәжірибесі. Зерттеудің жаңалығы жекелеген цифрлық құралдың тиімділігін қайта дәлелдеуде емес. Мұнда Қазақстандағы бастауыш сынып мұғалімдерінің нақты қолдану тәжірибесі, цифрлық даярлығы, техникалық мүмкіндігі және кездесетін кедергілері бір талдау аясында қарастырылады. Алынған нәтижелер мұғалімдерге арналған цифрлық даярлық курстарын нақтылауға, сабаққа бейімделген әдістемелік нұсқаулық әзірлеуге және мектептердегі техникалық қолдаудың әлсіз тұстарын анықтауға пайдаланылуы мүмкін.

Осы мақсатқа жету үшін мынадай **зерттеу сұрақтары** айқындалды:

1. Бастауыш сынып мұғалімдері оқушылардың кітап оқуға қызығушылығын арттыру мақсатында цифрлық технологияларды қаншалықты жиі қолданады?
2. Мұғалімдер цифрлық технологиялардың оқушылардың кітап оқуға қызығушылығын арттырудағы тиімділігін қалай бағалайды?
3. Цифрлық технологияларды қолдану жиілігі олардың қабылданатын тиімділігімен және мұғалімдердің цифрлық технологияларды қолдануға даярлығымен байланысты ма?
4. Цифрлық технологияларды қолдану жиілігі педагогикалық өтілі әртүрлі мұғалімдер арасында ерекшелене ме?
5. Мұғалімдер оқушылардың кітап оқуға қызығушылығын арттыру үшін цифрлық технологияларды қолдану барысында қандай кедергілерге кездеседі?

Гипотезалар

H1: Цифрлық технологияларды қолдану жиілігі мен мұғалімдердің олардың кітап оқуға қызығушылықты арттырудағы тиімділігін бағалауы арасында оң және статистикалық мәнді байланыс бар.

H2: Мұғалімдердің цифрлық технологияларды қолдануға даярлығы мен оларды қолдану жиілігі арасында оң және статистикалық мәнді байланыс бар.

H3: Цифрлық технологияларды қолдану жиілігі педагогикалық өтілі әртүрлі мұғалімдер арасында статистикалық мәнді ерекшеленеді.

Әдеби шолу

Бастауыш сыныпта электрондық кітап, оқу платформасы, мобильді қосымша немесе толықтырылған шындық құралы қолданылғаны сабақтың сапасы бірден артты дегенді білдірмейді. Бір құрал оқушыны мәтінге қайта оралтуы мүмкін, екіншісі назарын анимация мен батырмаларға бөліп жібереді. Айырмашылық құрылғыда емес, цифрлық мазмұнның оқу мақсатына қаншалықты дәл бағындырылғанында. Оқушының жасы, ақпаратты қабылдау ерекшелігі және мұғалімнің цифрлық-педагогикалық даярлығы да нәтижеге тікелей араласады [12]. Сондықтан технология жиі қолданылды деген көрсеткішті оқу сапасының дәлелі ретінде қабылдау қиын. Баланың экрандағы мәтінге қызыға қарауы әлі оқуға тартылды деген сөз емес. Тартылудың кемінде төрт қыры бар: тапсырманы аяқтауға ұмтылу және оқуға уақыт бөлу; мәтінді түсіну стратегияларын пайдалану; қызығушылық пен жағымды эмоция сезіну; оқығанын басқа балалармен талқылау. Alhamad (2024) жүргізген сапалық зерттеуде AR кітаптар осы төрт бағыттағы белсенділікті күшейткен [13]. Үшөлшемді кейіпкер, дыбыстық сүйемелдеу және интерактивті әрекет балаларды оқиғаға қайта үңілуге, сюжетті құрдастарымен талқылауға итермелеген. Бірақ нәтиже кез келген AR кітап пайдалы дегенге келмейді. Үшөлшемді бейне мәтіндегі оқиғаны түсіндіруге қызмет етпесе, ол оқу құралынан гөрі назар аудартатын эффект болып қалуы мүмкін. Цифрлық ресурстар қызығушылықпен ғана жұмыс істемейді. Dean (2021) қарастырған интервенцияларда компьютерлік бағдарламалар мен арнайы қосымшалар әріпті дыбысқа айналдыру, сөзді тану, оқу жылдамдығын арттыру және мәтінді түсіну дағдылары үшін қолданылған [14]. Oakley (2024) жүйелеген зерттеулерде қайталап оқу, сөздің дұрыс дыбысталу үлгісін тыңдау және автоматты кері байланыс беру пайдалы тетік ретінде аталады [15]. Әсердің көлемі барлық балада бірдей емес. Pasqualotto (2025) метаталдауында цифрлық

интервенциялар оқуында қиындығы бар балаларға орташа оң әсер, ал жалпы оқушылар тобына шағын оң әсер берген [16]. Декодтау мен мәтінді түсінуге тікелей бағытталған тапсырмалар екі топқа да көмектескен. Зейін мен жұмыс жадын жаттықтыратын жалпы когнитивтік бағдарламалардың пайдасы, негізінен, оқуында қиындық кездесетін балалардан байқалған. Яғни бір сыныпқа бірдей қосымша ұсыну ыңғайлы болғанымен, педагогикалық тұрғыдан дәл шешім болмауы мүмкін. Мультимедианың мәтінмен байланысы ерекше маңызды. Şimşek (2024) мектепке дейінгі жастағы 90 баланы қамтыған экспериментте AR форматындағы әңгімені оқыған топ оқығаны түсіну және қайта баяндау көрсеткіштері бойынша баспа кітаппен жұмыс істеген топтан жоғары нәтиже көрсеткенін анықтады [17]. Дыбыс, бейне және интерактивті элементтер сюжетті қайталап тұрған жоқ, баланың назарын кейіпкер әрекетіне, оқиғалардың ретіне және себеп-салдарлық байланыстарға бағыттаған. Дегенмен, 90 мектепке дейінгі баламен және шектеулі мәтіндер жиынтығымен алынған нәтижені бастауыш мектептің 1-4-сыныптарына тұтас көшіруге негіз аз. Жас ұлғайған сайын мәтін күрделенеді, ал оқу мақсаты оқығаны қайталап айтудан ақпаратты талдауға қарай өзгереді. Экран мен қағазды салыстырғанда да біржақты жауап шықпайды. Furenes (2021) метаталдауында балалардың мәтінді түсіну көрсеткіші жалпы есепте баспа форматында жоғары болған [1]. Айырмашылық электрондық нұсқа қағаз мәтіннің экранға көшірілген қарапайым нұсқасы болғанда күшейген. Ал сөздік, дыбыстау немесе анимация мәтіндегі сөз бен оқығаны түсіндіруге тікелей көмектескен жағдайда экранның кемшілігі азайған, кей көрсеткіштерде цифрлық формат алға шыққан. Мәселе «қағаз ба, экран ба?» деген таңдаудан тар емес. Негізгі сұрақ – экрандағы әр мүмкіндік оқушыға мәтінмен қандай әрекет жасатады? Қажетсіз анимация, белсенді сілтеме, дыбыс немесе ойын тапсырмасы жұмыс жадының бір бөлігін өзіне тартады. Оқушы мәтіндегі себеп-салдарлық байланысты қадағалаудың орнына құрылғыны басқаруға көшеді. Әсіресе экран ойын-сауық құралы ретінде қабылданса, мәтінді жылдам шолу және түсінбеген тұсын байқамай өту қаупі өседі. Қысқа мерзімді қызығушылық бар. Терең түсіну болмауы мүмкін. Сондықтан мультимедиалық элементтердің саны емес, сюжетті немесе ұғымды ашуға қажеттілігі бағалануы керек. Aleksieva (2025) зерттеуінде бастауыш сынып мұғалімдері сұхбат кезінде оқушы дербестігін, ынтымақтастықты және шығармашылықты қолдайтынын айтқан [18]. Сабақты бақылау басқа көрініс берген: технология көбіне ақпарат көрсетуге, дайын тапсырма ұсынуға және мұғалім басқаратын дәстүрлі жұмысты жалғастыруға пайдаланылған. Интерактивті тақта болғанымен, оқушының әрекеті мұғалім көрсеткен слайдты көруден аспауы ықтимал. Техникалық дағдының өзі технологиялық, педагогикалық және пәндік білімді біріктірмейді. Мұғалім құралды таңдағанда үш нәрсені қатар тексеруі қажет, оқу мақсаты, мәтіннің сипаты және оқушы орындайтын нақты әрекет. Жауапкершілікті тек мұғалімге жүктеу де мәселені жеңілдетіп көрсетеді. Timotheou (2023) талдаған еңбектерде мектептің цифрлық әлеуетіне интернет сапасы, құрылғы саны, мазмұнның қолжетімділігі, кәсіби қолдау, сабаққа дайындалатын уақыт, мектеп басшылығы және отбасының әлеуметтік-экономикалық жағдайы қатар ықпал еткені көрсетілген [12]. Мәселен, бір сыныпта 25 оқушыға 10 планшет берілуі немесе сабақ үстінде интернеттің үзілуі мұғалімнің жақсы жоспарын іске асырмай тастауы мүмкін. Құрылғы сатып алу – жұмыстың бір бөлігі ғана. Сапалы мазмұн, техникалық қызмет және мұғалімге берілетін әдістемелік қолдау қатар жүрмесе, цифрландыру жекелеген қосымшаларды қолданумен шектеледі. Зерттеулерден қарама-қайшы емес, шартқа тәуелді көрініс шығады. AR элементтері баланың эмоциялық қызығушылығын күшейтіп, мәтінге қайта оралуына түрткі бере алады; бірақ сюжетке қатысы жоқ эффект түсінуге кедергі жасайды. Цифрлық интервенциялар оқуында қиындығы бар балаларға көбірек көмектесуі мүмкін, алайда тапсырма баланың нақты оқу қажеттілігіне бағытталуы керек. Экран кей жағдайда қағаздан әлсіз, кей жағдайда пайдалырақ. Шешуші өлшем – технологияның қолданылу жиілігі емес, оқушыға қандай танымдық жұмыс орындататыны. Сол себепті цифрлық құралды қағаз кітаптың толық баламасы ретінде емес, нақты оқу қиындығын шешетін және мұғалім мақсатты түрде таңдайтын қосымша ресурс ретінде қарастырған дұрыс.

Зерттеу материалдары мен әдістемесі

Зерттеу бастауыш сынып мұғалімдерінің оқушылардың кітап оқуға қызығушылығын арттыруда цифрлық технологияларды қолдану ерекшеліктерін анықтауға бағытталды. Зерттеуде сандық, көлденең қималық сауалнама дизайны қолданылды. Бастапқыда сауалнамаға 156 мұғалім қатысты. Деректерді алдын ала тексеру барысында жеті респонденттің сауалнама сұрақтарының басым бөлігіне жауап бермегені анықталды. Ақпараттың толық болмауы статистикалық талдау нәтижелерінің сенімділігіне әсер етуі мүмкін болғандықтан, бос жауаптары көп осы сауалнамалар деректер жиынтығынан алынып тасталды. Нәтижесінде талдауға толық толтырылған 149 жауаптар енгізілді. Қатысушылар қолжетімді іріктеу әдісімен таңдалды. Қатысушылар педагогикалық өтіліне, сабақ беретін сыныбына және білім беру ұйымының орналасуына қатысты мәліметтер ұсынды. Ауылдық жерден қатысушылар саны өте аз болғандықтан, елді мекен түрі бойынша салыстырмалы статистикалық талдау жүргізілген жоқ. Негізгі деректерді жинау алдында сауалнама сұрақтарының түсініктілігін, мазмұндық сәйкестігін және жауап беру рәсімінің қолайлылығын бағалау мақсатында мұғалімдердің шағын тобымен пилоттық сынақ өткізілді. Пилоттық сынақ нәтижесінде сұрақтардың тұжырымдары мен жауап нұсқалары нақтыланып, түсініксіз немесе мағыналық тұрғыдан қайталанатын тармақтарға редакциялық түзетулер енгізілді. Пилоттық сынаққа қатысқан мұғалімдердің жауаптары негізгі статистикалық талдауға енгізілген жоқ.

Деректер 2026 жылдың ақпан-наурыз айларында Google Forms платформасында орналастырылған құрылымдалған сауалнама арқылы онлайн форматта жиналды. Сауалнама сұрақтары зерттеудің мақсаты мен міндеттеріне сәйкес автор тарапынан әзірленді. Сауалнаманың мазмұндық валидтілігі педагогика және цифрлық білім беру саласындағы екі тәуелсіз сарапшының бағалауы арқылы тексерілді. Сарапшылар сауалнама тұжырымдарының зерттеу мақсатына сәйкестігін, мазмұндық қамтылуын және түсініктілігін бағалады. Сауалнама бірнеше бөлімнен тұрды:

1. қатысушылардың әлеуметтік-демографиялық және кәсіби сипаттамалары;
2. цифрлық технологияларды қолдану жиілігін анықтайтын 7 тұжырым;
3. цифрлық технологиялардың оқушылардың кітап оқуға қызығушылығына ықпалын бағалайтын 5 тұжырым;
4. мұғалімдердің цифрлық технологияларды қолдануға даярлығы мен техникалық мүмкіндігін бағалайтын сұрақтар;
5. цифрлық технологияларды қолдануға кедергі келтіретін факторларды анықтайтын 4 тұжырым.

Қолдану жиілігі «ешқашан» деген 1 балдан «өте жиі» деген 5 балға дейінгі шкаламен бағаланды. Қабылданатын әсер, мұғалімнің даярлығы, техникалық мүмкіндік және кедергілер туралы тұжырымдар бес балдық Лайкерт шкаласымен өлшенді: 1 балл – «мүлдем келіспеймін», 5 балл – «толық келісемін». Балл жоғарылаған сайын өлшенетін көрсеткіштің деңгейі де жоғары деп қабылданды. Әр шкаланың интегралдық көрсеткіші жауаптардың арифметикалық орташа мәні арқылы есептелді. Толық емес сауалнамалар бірден алып тасталған жоқ. «Қолдану жиілігі» көрсеткішін есептеу үшін респондент жеті сұрақтың кемінде төртеуіне, «қабылданатын әсер» көрсеткіші үшін бес сұрақтың кемінде үшеуіне, ал «кедергілер» көрсеткіші үшін төрт сұрақтың кемінде екеуіне жауап беруі қажет болды. Осы шектен төмен жауап берілген жағдайда тиісті шкаланың интегралдық көрсеткіші есептелмеді.

Сауалнамаға қатысу ерікті болды. Қатысушыларға зерттеудің мақсаты, жауаптардың құпия сақталатыны және деректердің тек ғылыми мақсатта пайдаланылатыны алдын ала түсіндірілді. Аты-жөні, жеке сәйкестендіру нөмірі немесе респондентті тікелей анықтауға мүмкіндік беретін басқа мәліметтер жиналмады. Деректер IBM SPSS Statistics 29 бағдарламасында өңделді. Талдауға дейін жауаптардың толықтығы тексеріліп, толық бос және талдауға жарамсыз сауалнамалар деректер жиынтығынан шығарылды. Алдымен жиілік

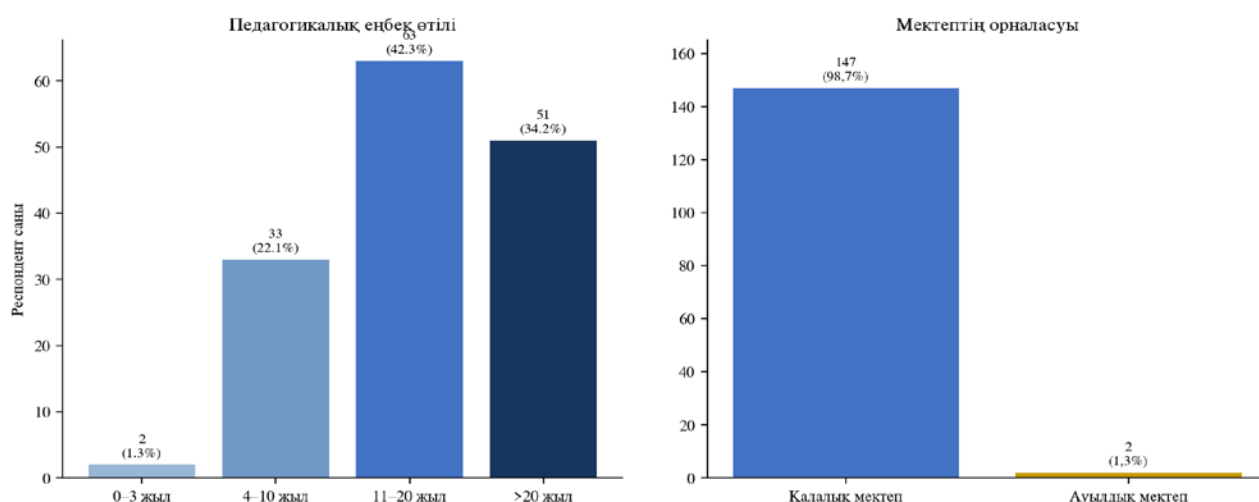
пен пайыздық үлес, арифметикалық орташа мән, стандарттық ауытқу, ең төменгі және ең жоғары мәндер есептелді. Шкалалардың ішкі сәйкестігі Кронбахтың альфа коэффициентімен бағаланды. Цифрлық технологияларды қолдану жиілігі, қабылданатын әсер және кедергілер шкалалары бойынша коэффициенттер 0,866 мен 0,962 аралығында болды. Мұндай мәндер аталған шкалалардағы тармақтардың өзара сәйкестігі жоғары екенін көрсетті. Айнымалылар бес балдық Лайкерт шкаласы арқылы өлшенгендіктен, олардың арасындағы байланыс Спирменнің рангтік корреляция коэффициентімен тексерілді. Талдауға цифрлық технологияларды қолдану жиілігі және оның қабылданатын әсермен, мұғалімнің даярлығымен, техникалық мүмкіндігімен және кедергілермен байланысы енгізілді. Корреляцияның болуы себеп-салдарлық ықпалды білдірмейді; ол көрсеткіштердің бірге өзгеру бағыты мен күшін ғана сипаттайды. Педагогикалық өтілі әртүрлі топтардың қолдану көрсеткіштерін салыстыру үшін параметрлік емес Краскел–Уоллис критерийі қолданылды. Барлық статистикалық тестілер екіжақты жүргізілді. Статистикалық мәнділік шегі $p < 0,05$ деңгейінде белгіленді.

Нәтижелер және талқылау

Зерттеу нәтижелері бойынша толық жауап берген 149 бастауыш сынып мұғалімінің деректері енгізіліп, бұл бастапқы іріктеменің 95,5%-ын құрады. Қорытынды іріктемедегі мұғалімдердің 147-сі (98,7%) қалалық, ал екеуі (1,3%) ауылдық мектептерде жұмыс істейді. Сондықтан зерттеу нәтижелері негізінен Қызылорда облысындағы қалалық мектеп мұғалімдерінің тәжірибесін сипаттайды. Респонденттердің басым бөлігі 11-20 жыл (42,3%) және 20 жылдан жоғары (34,2%) еңбек өтілі бар тәжірибелі педагогтер болды. Сыныптар бойынша бөлініс салыстырмалы түрде тең: 1-сыныпта 35, 2-сыныпта 36, 3-сыныпта 40 және 4-сыныпта 37 мұғалім сабақ береді; бір респондент сыныпты көрсетпеген.

Кесте 1 – Респонденттердің әлеуметтік-кәсіби сипаттамасы

Көрсеткіш	Санат	Саны
Педагогикалық өтіл	0–3 жыл	2
	4–10 жыл	33
	11–20 жыл	63
	20 жылдан жоғары	51
Сабақ беретін сынып	1-сынып	35
	2-сынып	36
	3-сынып	40
	4-сынып	37
	Жауап жоқ	1
Мектептің орналасуы	Қала	147
	Ауыл	2
<i>Ескертпе. Пайыздар толық іріктемеге (N=149) есептелді</i>		



1-сурет – Респонденттердің еңбек өтілі және мектептің орналасуы бойынша бөлінісі

Іріктеменің 98,7%-ын қалалық мектеп мұғалімдері құрағандықтан, төмендегі нәтижелер негізінен қалалық білім беру ортасының тәжірибесін сипаттайды. Ауыл және қала мектептерін статистикалық салыстыруға ауылдық топтың саны жеткіліксіз.

Сауалнамадағы үш көптармақты шкаланың ішкі сәйкестілігі Cronbach α коэффициентімен бағаланды. Қолдану жиілігі шкаласының сенімділігі өте жоғары ($\alpha=0,953$), ал қабылданатын әсер ($\alpha=0,855$) және кедергілер ($\alpha=0,851$) шкалаларының сенімділігі жоғары болды.

Кесте 2 – Негізгі көрсеткіштердің сипаттамалық статистикасы және шкала сенімділігі

Көрсеткіш	n	M	SD	Md	α
Цифрлық технологияларды қолдану жиілігі	149	2,74	1,11	2,43	0,953
Цифрлық технологиялардың қабылданатын әсері	149	3,87	0,66	4,00	0,855
Мұғалімнің цифрлық даярлығы	149	3,76	0,83	4,00	–
Мектептің техникалық мүмкіндігі	149	3,72	0,94	4,00	–
Қолдануға кедергі келтіретін факторлар	149	3,40	0,78	3,25	0,851

Ескертпе. M – орташа мән; SD – стандарттық ауытқу; Md – медиана; α – Cronbach коэффициенті. Барлық көрсеткіш 1–5 шкаласында өлшенді.

Нәтижелердің басты ерекшелігі – қабылданатын әсер мен нақты қолдану арасындағы алшақтық. Мұғалімдер цифрлық технологиялардың педагогикалық мүмкіндігін жоғары бағалаған – орташа көрсеткіш 3,87 балл. Бірақ сабақтағы нақты қолдану жиілігі небәрі 2,74 балл болып, бес балдық шкаланың теориялық ортасына да жетпеді. Демек, технологияны пайдалы деп санау оны жүйелі қолдануға әлі кепіл емес. Мәселе тек құралдың қолжетімділігінде емес. Мұғалімдер өз даярлығын 3,76, ал техникалық мүмкіндіктерді 3,72 баллмен бағалаған. Соған қарамастан, кедергілер көрсеткіші де төмен болған жоқ – 3,40 балл. Яғни мұғалімде құрылғы мен бастапқы дағды болғанымен, уақыт тапшылығы, әдістемелік қолдаудың аздығы немесе сабаққа лайық мазмұнның жетіспеуі технологияны тұрақты қолдануға бөгет болуы мүмкін. Жеті технология түрінің ешқайсысы 3 балдан асқан жоқ: орташа мәндер 2,62–2,97 аралығында қалды. Ең жиі қолданылғаны – электрондық

оқулықтар ($M=2,97$). Одан кейін анимациялар мен қысқа бейнелер ($M=2,81$), интерактивті тақтада мәтін көрсету ($M=2,74$) орналасқан. Тіпті көш бастаған электрондық оқулықтың өзі шкаланың орта деңгейіне жетпеген. Сондықтан мұғалімдердің цифрлық технологияға көзқарасы оң болғанымен, оны кітап оқуға қызығушылықты арттыратын күнделікті педагогикалық тәжірибеге айналдыру әлі толық іске аспаған.

Кесте 3 – Цифрлық технология түрлерін қолдану көрсеткіштері

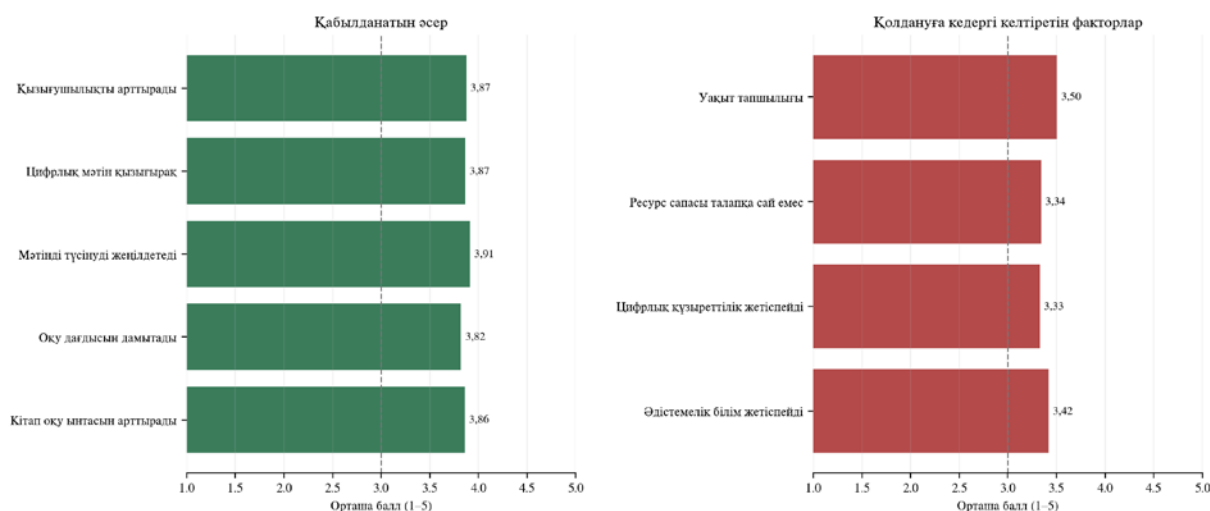
Технология түрі	n	M	SD	4-5 жауаптары, %
Электронды оқулықтар	149	2,97	1,19	21,5
Интерактивті тақтада мәтін	149	2,74	1,35	25,5
Аудиокітап	149	2,66	1,24	21,5
Анимация және қысқа бейне	149	2,81	1,24	23,5
Интернет платформалардағы тапсырма	149	2,66	1,26	21,5
Онлайн платформалардағы мәтін	149	2,62	1,24	20,8
Жасанды интеллект	149	2,68	1,27	22,1
<i>Ескертпе. 4-5 жауаптарының үлесі технологияны жиі немесе өте жиі қолданатын респонденттерді көрсетеді</i>				

Негізгі цифрлық құралды таңдау сұрағында интерактивті тақта бірінші орында болды: оны 74 респондент (49,7%) атаған. Арнайы білім платформаларын 21 мұғалім (14,1%), компьютерді 24 мұғалім (16,1%) таңдады. Бұл екі нәтиже бір-бірін толықтырады: интерактивті тақта негізгі құрал ретінде кең таралғанымен, оқу мәтіндерімен жұмыс істеуде оны тұрақты пайдалану жиілігі орташа деңгейге жетпеген ($M=2,74$).

Қабылданатын әсердің барлық тармағы жоғары бағаланды. Мұғалімдердің 72,5%-ы цифрлық технология мәтінді түсінуді жеңілдетеді деген пікірмен келісті немесе толық келісті; бұл тармақ ең жоғары орташа мәнге ие болды ($M=3,91$). Қызығушылықты арттыру және цифрлық мәтіннің тартымдылығы жөніндегі екі тармақта да жоғары жауаптардың үлесі 71,1% болды.

Кесте 4 – Қабылданатын әсер мен кедергілердің тармақтық көрсеткіштері

Блок	Тармақ	M	SD	4-5 жауаптары, %
Қабылданатын әсер	Қызығушылықты арттырады	3,87	0,82	71,1
Қабылданатын әсер	Цифрлық мәтін қызығырақ	3,87	0,86	71,1
Қабылданатын әсер	Мәтінді түсінуді жеңілдетеді	3,91	0,75	72,5
Қабылданатын әсер	Оқу дағдысын дамытады	3,82	0,86	65,8
Қабылданатын әсер	Кітап оқу ынтасын арттырады	3,86	0,87	66,4
Кедергі	Уақыт тапшылығы	3,50	0,89	53,0
Кедергі	Ресурс сапасы талапқа сай емес	3,34	0,86	43,0
Кедергі	Цифрлық құзыреттілік жетіспейді	3,33	0,99	45,0
Кедергі	Әдістемелік білім жетіспейді	3,42	0,99	46,3
<i>Ескертпе. Қабылданатын әсер тармақтарында 4-5 – келісуді; кедергі тармақтарында 4-5 – мәселенің бар екенін көрсетеді</i>				



2-сурет – Қабылданатын әсер мен цифрлық технологияларды қолдануға кедергі келтіретін факторлардың орташа мәндері

Кедергілер арасында уақыт тапшылығы жетекші орын алды ($M=3,50$): мұғалімдердің 53,0%-ы бұл мәселемен келісті немесе толық келісті. Әдістемелік білімнің жеткіліксіздігін 46,3%, цифрлық құзыреттіліктің жеткіліксіздігін 45,0%, ал ресурстар сапасының талапқа сай еместігін 43,0% белгіледі. Демек, кедергілер тек техникалық сипатта емес; уақытты ұйымдастыру мен әдістемелік қолдау да маңызды.

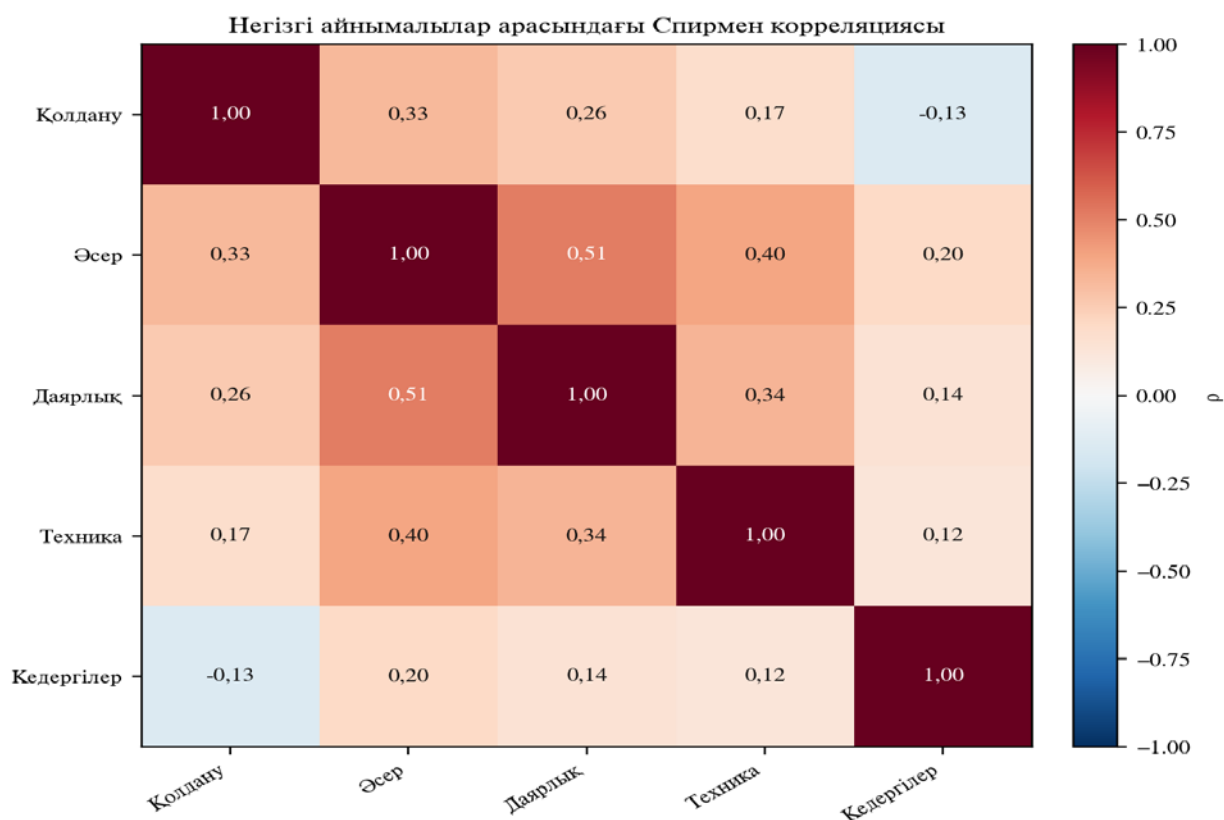
Шкала көрсеткіштерінің қалыпты үлестірім шарты орындалмағандықтан, байланыстар Spearman рангілік корреляциясы арқылы тексерілді. Бірнеше салыстырудан туындайтын жалған-оң нәтиже тәуекелін азайту үшін 10 корреляцияға Holm түзетуі қолданылды.

Кесте 5 – Негізгі айнымалылар арасындағы Spearman корреляциясы

Айнымалылар	n	ρ	p	Holm p	Қорытынды
Қолдану – Қабылданатын әсер	149	0,326	<0,001	<0,001	Мәнді
Қолдану – Даярлық	149	0,264	0,001	0,007	Мәнді
Қолдану – Техникалық мүмкіндік	149	0,174	0,033	0,134	Мәнді емес
Қолдану – Кедергілер	149	-0,134	0,102	0,257	Мәнді емес
Қабылданатын әсер – Даярлық	149	0,509	<0,001	<0,001	Мәнді
Қабылданатын әсер – Техникалық мүмкіндік	149	0,396	<0,001	<0,001	Мәнді
Қабылданатын әсер – Кедергілер	149	0,203	0,013	0,066	Мәнді емес
Даярлық – Техникалық мүмкіндік	149	0,338	<0,001	<0,001	Мәнді
Даярлық – Кедергілер	149	0,141	0,086	0,257	Мәнді емес
Техникалық мүмкіндік – Кедергілер	149	0,118	0,151	0,257	Мәнді емес

Ескертпе. Екіжақты статистикалық маңыздылық шегі $\alpha=0,05$. Негізгі интерпретация Holm түзетілген p мәніне негізделді.

Қолдану жиілігі қабылданатын әсермен әлсіз оң байланысты болды ($\rho=0,326$; Holm $p<0,001$), яғни технологияны жиірек қолданатын мұғалімдер оның әсерін де жоғары бағалаған. Қолдану мен даярлық арасында да әлсіз оң байланыс анықталды ($\rho=0,264$; Holm $p=0,007$). Ең күшті байланыс қабылданатын әсер мен мұғалім даярлығы арасында байқалды ($\rho=0,509$; Holm $p<0,001$). Сонымен қатар қабылданатын әсер техникалық мүмкіндікпен ($\rho=0,396$; Holm $p<0,001$), ал даярлық техникалық мүмкіндікпен ($\rho=0,338$; Holm $p<0,001$) оң байланысты болды.



3-сурет – Негізгі айнымалылардың корреляциялық матрицасы

Бастапқы p мәні бойынша қолдану мен техникалық мүмкіндік арасындағы байланыс мәнді көрінгенімен ($p=0,174$; $p=0,033$), Holm түзетуінен кейін статистикалық мәнділігін жоғалтты ($p=0,134$). Сондықтан бұл байланысты тұрақты нәтиже ретінде қарастыруға болмайды. Корреляциялар себеп-салдарлық қатынасты дәлелдемейді.

0-3 жылдық өтілі бар мұғалімдер саны өте аз болғандықтан ($n=2$), олар 4-10 жылдық топпен біріктірілді. Үш өтіл тобы Kruskal–Wallis тестімен салыстырылды. Бес негізгі көрсеткіштің ешқайсысында статистикалық мәнді айырмашылық анықталмады.

Кесте 6 – Негізгі көрсеткіштердің еңбек өтілі бойынша айырмашылығы

Көрсеткіш	Топ медианалары	Н	df	p	ε^2
Қолдану	0–10 жыл: 2,43; 11–20 жыл: 2,29; 20 жылдан жоғары: 2,43	0,182	2	0,913	0
Қабылданатын әсер	0–10 жыл: 4,00; 11–20 жыл: 4,00; 20 жылдан жоғары: 4,00	0,323	2	0,851	0
Даярлық	0–10 жыл: 4,00; 11–20 жыл: 4,00; 20 жылдан жоғары: 4,00	0,026	2	0,987	0
Техникалық мүмкіндік	0–10 жыл: 3,00; 11–20 жыл: 4,00; 20 жылдан жоғары: 4,00	3,713	2	0,156	0,012
Кедергілер	0–10 жыл: 3,00; 11–20 жыл: 3,25; 20 жылдан жоғары: 3,50	2,915	2	0,233	0,006

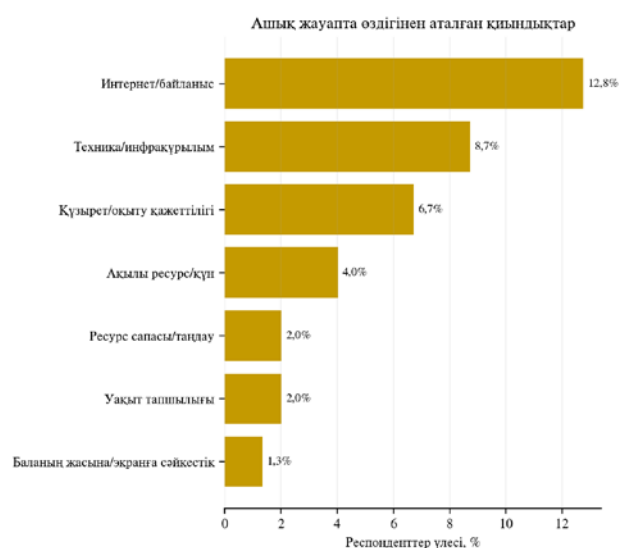
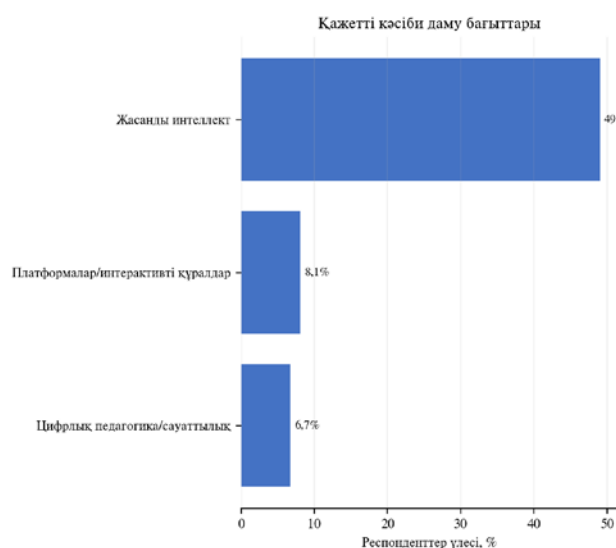
Ескертпе. Топтар: 0–10 жыл ($n=35$), 11–20 жыл ($n=63$), 20 жылдан жоғары ($n=51$). Барлық $p > 0,05$; ε^2 әсер мөлшері өте аз.

Қолдану жиілігі ($N=0,182$; $p=0,913$), қабылданатын әсер ($N=0,323$; $p=0,851$), даярлық ($N=0,026$; $p=0,987$), техникалық мүмкіндік ($N=3,713$; $p=0,156$) және кедергілер ($N=2,915$; $p=0,233$) бойынша топтық айырмашылықтар мәнді болған жоқ. Бұл іріктемеде цифрлық технологияны қолдану ерекшеліктерін еңбек өтілінен гөрі даярлық, әдістемелік қолдау немесе ұйымдық жағдайлар жақсырақ түсіндіруі ықтимал. Ашық жауаптар кең тақырыптық санаттарға кодталды. Бір жауап бірнеше санатқа енгізілуі мүмкін, сондықтан көпжауапты санаттардың пайыздары 100%-ға тең болмауы ықтимал.

Кесте 7 – Ашық жауаптардан анықталған негізгі тақырыптар

Сұрақ блогы	Санат	Саны
Өздігінен аталған қиындық	Қиындық жоқ	79
	Интернет байланыс	19
	Техника инфрақұрылым	13
	Құзырет және оқыту қажеттілігі	10
	Ақылы ресурс күн	6
Кәсіби даму қажеттілігі	Жасанды интеллект	73
	Платформалар және интерактивті құралдар	12
	Цифрлық педагогика сауаттылық	10
Цифрлық технологияны негізгі әдіс деп бағалау	Келіседі	99
	Келіспейді	33
	Ішінара бейтарап	9
	Нақты емес	8

Ескертпе. Қиындықтар мен курс қажеттіліктері көпжауапты кодтау арқылы есептелді. «Нақты емес/басқа» санаттары ықшамдық үшін кестеде көрсетілмеді



4-сурет – Ашық жауаптарда аталған кәсіби даму қажеттіліктері мен қиындықтар

Респонденттердің 53,0%-ы ашық сұрақта нақты қиындық атамады. Дегенмен қиындықты өздігінен көрсеткен жауаптарда интернет байланысы (12,8%), техника мен инфрақұрылым (8,7%) және құзыретті арттыру қажеттілігі (6,7%) басым болды. Жабық сұрақта уақыт тапшылығының жоғары бағалануы (53,0%) мен ашық сұрақта оның сирек өздігінен аталуы (2,0%) сұрақ форматына байланысты айырмашылықты көрсетеді: берілген нұсқаны бағалау мен мәселені өздігінен еске түсіру бірдей өлшем емес.

Кәсіби даму бағыттарының ішінде жасанды интеллект айқын басым болды: оны 73 мұғалім (49,0%) атады. Интерактивті платформаларға 8,1%, ал цифрлық педагогика мен

сауаттылыққа 6,7% сұраныс білдірілді. Сонымен қатар 99 респондент (66,4%) цифрлық технологияны кітап оқуға қызығушылықты арттырудың негізгі әдісі ретінде қабылдайтынын білдірді.

Зерттеу гипотезалары $\alpha=0,05$ статистикалық маңыздылық деңгейінде тексерілді. Н1 және Н2 гипотезалары үшін Spearman рангілік корреляциясы, ал Н3 гипотезасы үшін Kruskal–Wallis тесті қолданылды. Корреляциялық талдауда бірнеше салыстырудың ықпалын ескеру мақсатында Holm түзетілген р мәндері де назарға алынды.

Кесте 8 – Зерттеу гипотезаларын статистикалық тексеру қорытындысы

Гипотеза	Тексерілген байланыс және айырмашылық	Статистикалық нәтиже	Шешім
Н1	Қолдану жиілігі – тиімділікті бағалау	$\rho=0,326$; $p<0,001$; Holm $p<0,001$	Қолдау тапты
Н2	Мұғалім даярлығы – қолдану жиілігі	$\rho=0,264$; $p=0,001$; Holm $p=0,007$	Қолдау тапты
Н3	Қолдану жиілігі – педагогикалық өтіл топтары	$H(2)=0,182$; $p=0,913$	Қолдау таппады

Ескертпе. N=149. ρ – Spearman корреляция коэффициенті; H – Kruskal–Wallis статистикасы. Н3 үшін 0–3 және 4–10 жыл топтары біріктірілді: 0–10 жыл (n=35), 11–20 жыл (n=63), 20 жылдан жоғары (n=51)

Н1 гипотезасы қолдау тапты. Цифрлық технологияларды қолдану жиілігі мен олардың оқушылардың кітап оқуға қызығушылығын арттырудағы тиімділігін бағалау арасында әлсіз, бірақ статистикалық мәнді оң байланыс анықталды ($\rho=0,326$; $p<0,001$; Holm $p<0,001$). Яғни цифрлық құралдарды жиірек қолданатын мұғалімдер олардың ықтимал педагогикалық әсерін де жоғары бағалаған. Дегенмен байланыстың әлсіз болуы тиімділікті бағалауға қолдану жиілігінен басқа факторлардың да ықпал ететінін көрсетеді.

Н2 гипотезасы қолдау тапты. Мұғалімдердің цифрлық технологияларды қолдануға даярлығы мен оларды қолдану жиілігі арасында әлсіз оң және статистикалық мәнді байланыс байқалды ($\rho=0,264$; $p=0,001$; Holm $p=0,007$). Демек, өзін цифрлық құралдарды қолдануға даяр деп бағалаған мұғалімдер оларды сабақта салыстырмалы түрде жиірек пайдаланған. Алайда бұл нәтиже даярлықтың қолдану жиілігін тікелей арттыратынын дәлелдемейді, өйткені корреляциялық талдау себеп-салдарлық қатынасты анықтамайды.

Н3 гипотезасы қолдау таппады. Цифрлық технологияларды қолдану жиілігі педагогикалық өтіл топтары бойынша статистикалық мәнді ерекшеленген жоқ ($H(2)=0,182$; $p=0,913$; $\varepsilon^2=0,000$). Осы іріктеме шегінде еңбек өтілі цифрлық технологияларды қолдану жиілігін түсіндіретін маңызды фактор болған жоқ. Бұл қорытынды еңбек өтілінің ықпалы мүлде жоқ дегенді білдірмейді; ең аз өтілі бар бастапқы топтың шағын болуы ($n=2$) салдарынан ол 4–10 жыл тобымен біріктірілді.

Жалпы алғанда, ұсынылған үш гипотезаның екеуі қолдау тапты: цифрлық технологияларды қолдану жиілігі олардың қабылданатын тиімділігімен және мұғалім даярлығымен оң байланысты болды. Педагогикалық өтіл бойынша күтілген топаралық айырмашылық расталмады. Сондықтан мұғалімдердің цифрлық технологияны қолдануын түсіндіруде еңбек өтілінен гөрі даярлық, әдістемелік қолдау және ұйымдастырушылық жағдайларды қарастыру негіздірек. Зерттеу нәтижелері мұғалімдердің цифрлық технологиялардың оқуға қызығушылықты арттыру мүмкіндігін жоғары бағалайтынын, бірақ бұл құралдарды нақты қолдану жиілігінің салыстырмалы түрде төмен екенін көрсетті. Үш зерттеу гипотезасының екеуі қолдау тапты: қолдану жиілігі мұғалім даярлығымен және қабылданатын әсермен оң байланысты болды; ал педагогикалық өтіл бойынша мәнді айырмашылық анықталмады. Осыған байланысты цифрлық құралдарды енгізу саясаты тек техника сатып алумен шектелмеуі тиіс; педагогтердің уақытын, әдістемелік даярлығын,

сапалы ресурстарға қолжетімділігін және тәжірибелік кәсіби дамуын қатар қамтуы қажет. Нәтижелерді түсіндіруде үш шектеу ескерілуі тиіс. Біріншіден, зерттеу көлденең қималы және өзіндік есепке негізделген, сондықтан себеп-салдарлық қорытынды жасауға болмайды. Екіншіден, қабылданатын әсер оқушылардың оқу қызығушылығы немесе оқу жетістігінің объективті өлшемі емес. Үшіншіден, іріктеменің қалалық мектептерге шоғырлануы нәтижелерді ауылдық мектептерге және Қазақстан мұғалімдерінің толық жиынтығына тікелей жалпылауды шектейді.

Қорытынды

Ұсынылған үш гипотезаның екеуі деректермен расталды. Цифрлық технологияларды жиі қолданатын мұғалімдер олардың оқушылардың кітап оқуға қызығушылығын арттыру мүмкіндігін де жоғары бағалаған. Қолдану жиілігі мұғалімнің даярлық деңгейімен де оң және статистикалық мәнді байланыс көрсетті. Демек, құралдың сабаққа енуі мұғалімнің оған қызығуымен ғана емес, оны оқу мақсатына сай пайдалана алуымен байланысты. Үшінші гипотеза расталмады: педагогикалық өтілі әртүрлі мұғалімдердің технологияны қолдану жиілігінде статистикалық мәнді айырмашылық табылған жоқ. Сондықтан цифрлық құралдарды жас мамандар жиі, ал тәжірибелі педагогтер сирек қолданады деп кесіп айтуға негіз жоқ. Бірақ мұны технология барлық буын мұғалімдеріне бірдей әсер етеді деп түсіндіру де асығыс болар еді. Зерттеу еңбек өтілі бойынша айырмашылықты ғана тексерді; мұғалімнің жасы мен технологияның «теңестіруші» әсері жеке өлшенген жоқ. Нәтижеде жиіліктен гөрі сабақтағы әрекет маңыздырақ болды. Экранға мәтінді шығару оқушыны кітап оқуға өздігінен тартпайды. Аудио, бейне және мәтін бір тапсырмада тоғысып, оқушыны салыстыруға, болжауға, қайта баяндауға немесе талқылауға жетелегенде ғана цифрлық құралдың танымдық қызметі күшейеді. Мұғалімдердің 49%-ы жасанды интеллект құралдарына қызығушылық білдірген. Сұраныс бар. Алайда респонденттердің 53%-ы уақыт тапшылығын, 46,3%-ы әдістемелік білімнің аздығын кедергі ретінде атаған. Яғни мәселе технологияға қызықпау емес: мұғалімге оны меңгеруге уақыт пен нақты әдістемелік қолдау жетіспейді. Осы екі мәселе шешілмесе, жасанды интеллект те, басқа цифрлық ресурстар да сабақта мазмұнды оқу әрекетінен гөрі ақпарат көрсететін қосымша құрал деңгейінде қалуы мүмкін.

Зерттеу нәтижелеріне сүйене отырып, мынадай практикалық ұсыныстар беріледі:

- Мұғалімдердің жұмыс жүктемесін оңтайландыру мақсатында цифрлық ресурстарды оқу бағдарламасына жүйелі түрде интеграциялау;
- Кәсіби даму курстарының мазмұнында жасанды интеллект құралдары мен интерактивті платформаларды әдістемелік тұрғыдан тиімді қолдануға басымдық беру;
- Техникалық инфрақұрылымды нығайтумен қатар, ресурстардың сапасына қатысты кедергілерді (43%) жою үшін сапалы қазақ тілді цифрлық контентті дамыту.

Зерттеудің негізгі шектеуі ретінде іріктеменің негізінен қалалық мұғалімдерден (98,7%) тұруын атап өткен жөн. Болашақта цифрлық теңсіздік мәселелерін тереңірек зерделеу үшін ауылдық жерлердегі білім беру ұйымдарын қамтитын кеңейтілген зерттеулер жүргізу қажет.

Әдебиеттер тізімі

1. Furenes M. I., Kucirkova N., Bus A. G. A Comparison of Children's Reading on Paper Versus Screen: A Meta-Analysis // Rev. Educ. Res. 2021. T. 91, № 4. С. 483–517. <https://doi.org/10.3102/0034654321998074>
2. Morgan P. L., Fuchs D. Is there a bidirectional relationship between children's reading skills and reading motivation? // Except. Child. SAGE Publications Sage CA: Los Angeles, CA, 2007. T. 73, № 2. С. 165–183. <https://doi.org/10.1177/001440290707300203>

3. Wigfield A., Guthrie J. T. Engagement and motivation in reading // *Handb. Read. Res.* 2000. T. 3, № 2000. C. 406.
4. Morrow L. M. The impact of a literature-based program on literacy achievement, use of literature, and attitudes of children from minority backgrounds // *Read. Res. Q. JSTOR*, 1992. C. 251–275.
5. McKenna M. C., Kear D. J. Measuring attitude toward reading: A new tool for teachers // *Read. Teach. JSTOR*, 1990. T. 43, № 9. C. 626–639.
6. Morrow L. M., Tracey D. H. *Lenses on reading: An introduction to theories and models*. Guilford Press, 2012.
7. Eutsler L. и др. The influence of mobile technologies on preschool and elementary children's literacy achievement: a systematic review spanning 2007–2019: L. Eutsler et al. // *Educ. Technol. Res. Dev.* Springer, 2020. T. 68, № 4. C. 1739–1768. <https://doi.org/10.1007/s11423-020-09786-1>
8. Takacs Z. K., Swart E. K., Bus A. G. Benefits and Pitfalls of Multimedia and Interactive Features in Technology-Enhanced Storybooks: A Meta-Analysis // *Rev. Educ. Res.* 2015. T. 85, № 4. C. 698–739. <https://doi.org/10.3102/0034654314566989>
9. Kao G. Y.-M. и др. The effects of high/low interactive electronic storybooks on elementary school students' reading motivation, story comprehension and chromatics concepts // *Comput. Educ. Elsevier*, 2016. T. 100. C. 56–70. <https://doi.org/10.1016/j.compedu.2016.04.013>
10. Salmerón L. и др. Relation between digital tool practices in the language arts classroom and reading comprehension scores // *Read. Writ.* 2023. T. 36, № 1. C. 175–194. <https://doi.org/10.1007/s11145-022-10295-1>
11. Fälth L., Selenius H. Primary school teachers' use and perception of digital technology in early reading and writing education in inclusive settings // *Disabil. Rehabil. Assist. Technol.* 2024. T. 19, № 3. C. 790–799. <https://doi.org/10.1080/17483107.2022.2125089>
12. Timotheou S. и др. Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review // *Educ. Inf. Technol.* 2023. T. 28, № 6. C. 6695–6726. <https://doi.org/10.1007/s10639-022-11431-8>
13. Alhamad K., Manches A., McGeown S. Augmented reality books: in-depth insights into children's reading engagement // *Front. Psychol.* 2024. T. 15. C. 1423163. <https://doi.org/10.3389/fpsyg.2024.1423163>
14. Dean J., Pascoe M., Le Roux J. Information and communication technology reading interventions: A scoping review // *Read. Writ.* 2021. T. 12, № 1. <https://doi.org/10.4102/rw.v12i1.294>
15. Oakley G. A Scoping Review of Research on the Use of Digital Technologies for Teaching Reading Fluency // *Educ. Sci.* 2024. T. 14, № 6. C. 633. <https://doi.org/10.3390/educsci14060633>
16. Pasqualotto A. и др. Digital Tools for Reading Success: Meta-analyses of Digital Interventions // *J. Cogn. Enhanc.* 2025. T. 9, № 4. C. 455–479. <https://doi.org/10.1007/s41465-025-00336-2>
17. Şimşek E. E. The effect of augmented reality storybooks on the story comprehension and retelling of preschool children // *Front. Psychol.* 2024. T. 15. C. 1459264. <https://doi.org/10.3389/fpsyg.2024.1459264>
18. Aleksieva L., Racheva V., Peytcheva-Forsyth R. Talking Tech, Teaching with Tech: How Primary Teachers Implement Digital Technologies in Practice // *Informatics.* 2025. T. 12, № 3. C. 99. <https://doi.org/10.3390/informatics12030099>

R. Kaliyeva, G. Yernazarova, M. Koishibayev, K. Kenzhebayeva

USING DIGITAL TECHNOLOGIES TO ENHANCE PRIMARY SCHOOL STUDENTS' INTEREST IN READING: TEACHERS' EXPERIENCES AND PERCEPTIONS

The presence of digital tools in the classroom does not, by itself, address declining interest in reading among primary school students. Even when interactive whiteboards and electronic textbooks are available, an important question remains: how often and in what ways do teachers use these tools specifically to strengthen students' reading motivation? This study, conducted in the Kyzylorda region, examined this gap. Initially, 156 primary school teachers participated in the quantitative survey. During data screening, seven responses that did not meet the inclusion criteria were excluded, leaving data from 149 teachers for the final analysis. The questionnaire covered not only the frequency of digital technology use, but also teachers' perceptions of its impact, their digital readiness, available technical resources, and barriers encountered in

everyday classroom practice. Data were analyzed using IBM SPSS Statistics 29. Spearman's rank correlation was used to examine relationships between variables, while the Kruskal–Wallis test was applied to compare differences across teacher groups. The findings revealed a notable discrepancy: teachers generally rated the pedagogical potential of digital tools highly, yet did not use them consistently in classroom practice. Electronic textbooks and interactive whiteboards were among the most frequently used tools. However, the availability of a device does not necessarily mean that its multimodal possibilities—including the integration of text, audio, video, and interactive tasks within a single learning activity—are being fully used. The issue, then, is not simply a matter of providing more technology. Responses from 149 teachers point to a need for practical methodological support that connects the use of digital tools with students' reading motivation. Improving teachers' digital readiness, providing classroom-oriented guidance, and addressing barriers to technology use therefore remain important priorities for primary education practice in Kazakhstan.

Keywords: primary school, interest in reading, digital technologies, reading motivation, teacher readiness, electronic textbooks, interactive learning.

Р.С. Калиева, Г.Ш.Ерназарова, М.Н.Койшыбаев, К.С.Кенжебаева

ИСПОЛЬЗОВАНИЕ ЦИФРОВЫХ ТЕХНОЛОГИЙ ДЛЯ ПОВЫШЕНИЯ ИНТЕРЕСА МЛАДШИХ ШКОЛЬНИКОВ К ЧТЕНИЮ: ОПЫТ И ОЦЕНКИ УЧИТЕЛЕЙ

Снижение интереса младших школьников к чтению нельзя решить одним лишь наличием цифровых средств в классе. Даже если установлена интерактивная доска и доступны электронные учебники, остается открытым вопрос, насколько часто и каким образом учителя используют их именно для повышения читательской мотивации. Исследование, проведенное в Кызылординской области, было направлено на уточнение этого разрыва. Первоначально в количественном опросе приняли участие 156 учителей начальных классов. При проверке данных 7 ответов, не соответствовавших установленным требованиям, были исключены, и в итоговый анализ вошли данные 149 педагогов. Анкета охватывала не только частоту использования цифровых технологий, но и оценку учителями их влияния, уровень цифровой готовности, технические возможности и трудности, возникающие в повседневной работе на уроке. Обработка данных проводилась в программе IBM SPSS Statistics 29. Для проверки связей между переменными использовалась ранговая корреляция Спирмена, а для сравнения различий между группами учителей – критерий Краскела–Уоллиса. Результаты выявили определенное противоречие: учителя высоко оценивают педагогический потенциал цифровых средств, однако используют их на уроках не всегда регулярно. Наиболее часто применяются электронные учебники и интерактивные доски. Вместе с тем наличие устройства еще не означает, что его мультимодальные возможности, включая объединение текста, звука, видео и интерактивных заданий в рамках одной учебной деятельности, используются в полной мере. Проблема не сводится к увеличению количества техники. Ответы 149 учителей указывают на необходимость конкретной методической поддержки, которая помогала бы связывать использование цифровых средств с повышением мотивации к чтению. В связи с этим совершенствование цифровой готовности педагогов, разработка практических рекомендаций, адаптированных к уроку, а также отдельное рассмотрение факторов, препятствующих использованию технологий, остаются важными задачами для практики отечественного начального образования.

Ключевые слова: начальная школа, интерес к чтению, цифровые технологии, учебная мотивация, готовность учителей, электронные учебники, интерактивное обучение.

References

1. Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research*, 91(4), 483–517. <https://doi.org/10.3102/0034654321998074>
2. Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children*, 73(2), 165–183. <https://doi.org/10.1177/001440290707300203>

3. Wigfield, A., & Guthrie, J. T. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 403–422). Lawrence Erlbaum Associates.
4. Morrow, L. M. (1992). The impact of a literature-based program on literacy achievement, use of literature, and attitudes of children from minority backgrounds. *Reading Research Quarterly*, 27(3), 250–275.
5. McKenna, M. C., & Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. *The Reading Teacher*, 43(8), 626–639.
6. Morrow, L. M., & Tracey, D. H. (2012). *Lenses on reading: An introduction to theories and models* (2nd ed.). Guilford Press.
7. Eutsler, L., Mitchell, C., Stamm, B., & Vehabovic, N. (2020). The influence of mobile technologies on preschool and elementary children's literacy achievement: A systematic review spanning 2007–2019. *Educational Technology Research and Development*, 68(4), 1739–1768. <https://doi.org/10.1007/s11423-020-09786-1>
8. Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research*, 85(4), 698–739. <https://doi.org/10.3102/0034654314566989>
9. Kao, G. Y.-M., Tsai, C.-C., Liu, C.-Y., & Yang, C.-H. (2016). The effects of high/low interactive electronic storybooks on elementary school students' reading motivation, story comprehension and chromatics concepts. *Computers & Education*, 100, 56–70. <https://doi.org/10.1016/j.compedu.2016.04.013>
10. Salmerón, L., Vargas, C., Delgado, P., & Baron, N. S. (2023). Relation between digital tool practices in the language arts classroom and reading comprehension scores. *Reading and Writing*, 36(1), 175–194. <https://doi.org/10.1007/s11145-022-10295-1>
11. Fälth, L., & Selenius, H. (2024). Primary school teachers' use and perception of digital technology in early reading and writing education in inclusive settings. *Disability and Rehabilitation: Assistive Technology*, 19(3), 790–799. <https://doi.org/10.1080/17483107.2022.2125089>
12. Timotheou, S., Miliou, O., Dimitriadis, Y., Soboleva, E. O., Srisawasdi, N., Cukurova, M., & Ioannou, A. (2023). Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. *Education and Information Technologies*, 28(6), 6695–6726. <https://doi.org/10.1007/s10639-022-11431-8>
13. Alhamad, K., Manches, A., & McGeown, S. (2024). Augmented reality books: In-depth insights into children's reading engagement. *Frontiers in Psychology*, 15, Article 1423163. <https://doi.org/10.3389/fpsyg.2024.1423163>
14. Dean, J., Pascoe, M., & Le Roux, J. (2021). Information and communication technology reading interventions: A scoping review. *Reading & Writing*, 12(1), Article a294. <https://doi.org/10.4102/rw.v12i1.294>
15. Oakley, G. (2024). A scoping review of research on the use of digital technologies for teaching reading fluency. *Education Sciences*, 14(6), Article 633. <https://doi.org/10.3390/educsci14060633>
16. Pasqualotto, A., Altarelli, I., Dehaene, S., & Bavelier, D. (2025). Digital tools for reading success: Meta-analyses of digital interventions. *Journal of Cognitive Enhancement*, 9(4), 455–479. <https://doi.org/10.1007/s41465-025-00336-2>
17. Şimşek, E. E. (2024). The effect of augmented reality storybooks on the story comprehension and retelling of preschool children. *Frontiers in Psychology*, 15, Article 1459264. <https://doi.org/10.3389/fpsyg.2024.1459264>
18. Aleksieva, L., Racheva, V., & Peytcheva-Forsyth, R. (2025). Talking tech, teaching with tech: How primary teachers implement digital technologies in practice. *Informatics*, 12(3), Article 99. <https://doi.org/10.3390/informatics12030099>

Авторлар туралы мәлімет:

Қалиева Раушангүл Сарсенәліқызы – «Бастауыш оқыту педагогикасы мен әдістемесі» білім беру бағдарламасының докторанты, Қорқыт Ата атындағы Қызылорда университеті, Қ.Сексенбаев көшесі 24-үй, e-mail: raushankalieva22@gmail.com

Ерназарова Гүлбаршын Шаршыбекқызы – филология ғылымдарының кандидаты, «Педагогика, психология және бастауыш оқыту әдістемесі» білім беру бағдарламаларының аға

оқытушысы, Қорқыт Ата атындағы Қызылорда университеті, С.Сейфуллин көшесі 46-үй, e-mail: guliya-09@mail.ru

Койшибаев Махамбетжан Нахипбекович – филология ғылымдарының кандидаты, «Педагогика, психология және бастауыш оқыту әдістемесі» білім беру бағдарламаларының аға оқытушысы, Қорқыт Ата атындағы Қызылорда университеті, Ә.Құдабаев көшесі 24 А., e-mail: mahat75.62@mail.ru

Кенжебаева Кундыз Сериковна (автор-корреспондент) – философия докторы (PhD), «Педагогика, психология және бастауыш оқыту әдістемесі» білім беру бағдарламасының аға оқытушысы, Қорқыт Ата атындағы Қызылорда университеті, Қойсары батыр көшесі 61-үй, e-mail: kundiz_27_87@mail.ru

Сведения об авторе:

Калиева Раушангуль Сарсеналиевна – докторант образовательной программы «Педагогика и методика начального обучения», Кызылординский университет имени Коркыт Ата, ул. К.Сексенбаева 24д., e-mail: raushankalieva22@gmail.com

Ерназарова Гульбаршын Шаршыбековна – кандидат филологических наук, старший преподаватель образовательных программ «Педагогика, психология и методика начального обучения», Кызылординский университет имени Коркыт Ата, ул. С.Сейфуллина 46д., e-mail: guliya-09@mail.ru

Койшибаев Махамбетжан Нахипбекович – кандидат филологических наук, старший преподаватель образовательных программ «Педагогика, психология и методика начального обучения», Кызылординский университет имени Коркыт Ата, ул. А.Кудабая 24а., e-mail: mahat75.62@mail.ru

Кенжебаева Кундыз Сериковна (автор-корреспондент) – доктор философии (PhD), старший преподаватель образовательных программ «Педагогика, психология и методика начального обучения», Кызылординский университет имени Коркыт Ата, ул. Койсары батыра 61д., e-mail: kundiz_27_87@mail.ru

Information about the author:

Kaliyeva Raushangul – doctoral student in the Educational Programme "Pedagogy and Methods of Primary Education", Korkyt Ata Kyzylorda University, K.Seksenbayeva St., 24, e-mail: raushankalieva22@gmail.com

Yernazarova Gulbarshyn – Candidate of Philological Sciences, senior lecturer of the Educational Program "Pedagogy, psychology and methods of primary education", Korkyt Ata Kyzylorda University, S.Seifullina str., 46, e-mail: guliya-09@mail.ru

Koishibayev Makhambetzhon – Candidate of Philological Sciences, senior lecturer of the Educational Program "Pedagogy, psychology and methods of primary education", Korkyt Ata Kyzylorda University, A.Kudabaev St., 24, e-mail: mahat75.62@mail.ru

Kenzhebayeva Kundyzy (corresponding author) – Doctor of Philosophy (PhD), senior lecturer of the educational programs "Pedagogy, psychology and methods of primary education", Korkyt Ata Kyzylorda University, Koisary batyr St., 61, e-mail: kundiz_27_87@mail.ru

Б.Б. Топоева-Ставинская¹, Г.С. Ерсултанова²¹Государственное учреждение «Кыргызтест», г. Бишкек, Кыргызская Республика²РГП на ПХВ «Национальный центр тестирования» МНВО РК, г. Астана, Казахстан

e-mail: bbstavinskaya@gmail.com

ORCID 0009-0004-8603-4929, ORCID 0000-0001-9503-7217

КОНЦЕПТУАЛЬНО-МЕТОДИЧЕСКИЕ ОСНОВЫ РАЗРАБОТКИ УРОВНЯ C2 В НАЦИОНАЛЬНЫХ СИСТЕМАХ ЯЗЫКОВОГО ТЕСТИРОВАНИЯ «КЫРГЫЗТЕСТ» И КАЗТЕСТ: СРАВНИТЕЛЬНАЯ МОДЕЛЬ ПЯТИ- И ЧЕТЫРЕХКОМПОНЕНТНОЙ АРХИТЕКТУРЫ

Актуальность исследования обусловлена необходимостью расширения национальных шкал языкового оценивания Кыргызстана и Казахстана до высшего уровня C2 при сохранении сопоставимости с международными подходами и национальной спецификой языков. Цель работы – обосновать общую концептуальную рамку C2 для систем «Кыргызтест» и КАЗТЕСТ и рассмотреть две операциональные модели: пятикомпонентную для «Кыргызтест» и четырехкомпонентную для КАЗТЕСТ. Методология включает дескрипторный анализ CEFR и CEFR Companion Volume, сопоставление действующих форматов Cambridge C2 Proficiency и Goethe-Zertifikat C2, анализ реформы Cambridge Proficiency, в результате которой пятикомпонентный экзамен был преобразован в четырехкомпонентный путем объединения Reading и Use of English, а также конструктивное проектирование на основе действующих архитектур двух национальных систем. Показано, что C2 должен отличаться от C1 не количественным усложнением материала, а качеством языковой деятельности: глубиной понимания, интерпретацией имплицитных смыслов, точностью выбора средств, регистровой гибкостью, интеграцией источников, нюансированной аргументацией и управлением взаимодействием. Формат сертификационного тестирования КАЗТЕСТ включает четыре блока: аудирование, чтение, письмо и говорение. Оценивание лексико-грамматической компетенции осуществляется посредством интегрированных заданий, включенных в блок «Чтение». «Кыргызтест» при этом сохраняет самостоятельный субтест «Лексика. Грамматика» как часть пятикомпонентной модели. Сделан вывод, что обе архитектуры могут обеспечивать валидную интерпретацию C2 при наличии спецификаций, многоэтапного пилотирования, психометрической проверки, стандартизации оценивания и процедуры установления порогов.

Ключевые слова: CEFR, C2, языковое тестирование, «Кыргызтест», КАЗТЕСТ, Национальная система оценивания, валидность, спецификация теста, психометрическая валидация, четырехкомпонентная модель.

Введение

Развитие национальных систем языкового тестирования неизбежно ставит вопрос о верхней границе уровневой шкалы. Для систем «Кыргызтест» и КАЗТЕСТ эта задача приобретает практическое значение в связи с необходимостью описывать и измерять владение государственными языками не только на функционально достаточных уровнях, но и на уровне профессионально, академически и общественно сложной коммуникации. CEFR определяет C2 как высший уровень своей шестиступенчатой шкалы, однако сама рамка не задает готового экзамена: национальная система должна самостоятельно операционализировать дескрипторы, определить конструкты, разработать типы заданий и представить доказательства валидности интерпретации результатов [1-3].

В первоначальной модели настоящего исследования C2 проектировался применительно к «Кыргызтест» как к пятикомпонентной системе с субтестами «Лексика. Грамматика», «Чтение», «Аудирование», «Письмо» и «Говорение». Практическая работа по взаимодействию «Кыргызтест» и КАЗТЕСТ показала необходимость расширить рамку

исследования и рассмотреть, каким образом одни и те же качественные признаки C2 могут быть реализованы в архитектуре, предполагаемой для сертификационного КАЗТЕСТ, где отдельного лексико-грамматического блока нет. Действующий сертификационный КАЗТЕСТ включает четыре блока – аудирование, чтение, письмо и говорение [10]. Настоящее исследование относится к сертификационному КАЗТЕСТ в его действующей редакции (по состоянию на 2026 г.); ранее структура КАЗТЕСТ включала блок «Лексико-грамматическая структура». Следовательно, научная проблема состоит не в выборе «правильного» числа субтестов, а в установлении того, обеспечивает ли каждая архитектура достаточное и недублирующее покрытие конструкта C2.

Такой подход согласуется с международной практикой. Cambridge C2 Proficiency в 2013 г. перешел от пяти экзаменационных компонентов к четырем: самостоятельные Reading и Use of English были объединены в единый компонент Reading and Use of English. Разработчики объясняли реформу частичным перекрытием измерительных фокусов и стремлением повысить эффективность экзамена без снижения охвата конструкта [7]. В современной версии экзамен состоит из четырех компонентов – Reading and Use of English, Writing, Listening и Speaking [6, 8]. Goethe-Zertifikat C2 также организован в четырех модулях: Lesen, Hören, Schreiben и Sprechen [9]. Эти примеры особенно важны для КАЗТЕСТ: они показывают, что лексико-грамматическая точность может измеряться не только отдельным субтестом, но и как распределенный конструкт внутри чтения, письма, говорения и, частично, аудирования.

При этом международный опыт нельзя переносить механически. Кыргызский и казахский языки имеют собственные структурные, функционально-стилевые и социолингвистические характеристики, а национальные системы обладают различной историей и организационной логикой. Поэтому задача заключается в сохранении общего дескрипторного ядра C2 при двух вариантах операционализации: пятикомпонентном для «Кыргызтест» и четырехкомпонентном для КАЗТЕСТ.

Цель исследования – разработать концептуально-методическую основу уровня C2 для двух национальных систем языкового оценивания и обосновать способы сохранения полноты конструкта при различной архитектуре экзамена. Для достижения цели решались следующие задачи: 1) выделить универсальные качественные признаки C2 на основе CEFR; 2) сопоставить международные четырехкомпонентные модели с национальными системами; 3) определить конструкт каждого субтеста/блока «Кыргызтест» и КАЗТЕСТ; 4) предложить модель интеграции лексико-грамматического контроля в четырехкомпонентный КАЗТЕСТ; 5) сформулировать требования к апробации, психометрической валидации и стандарту C2.

Научная новизна работы состоит в том, что уровень C2 рассматривается не как единая неизменная экзаменационная форма, а как общий конструкт, допускающий разные архитектурные решения при сохранении измерительной полноты. Практическая значимость заключается в предложении рабочей спецификационной рамки, которая может быть использована при разработке экспериментальных банков заданий уровня C2 для кыргызского и казахского языков.

Материалы и методы исследования

Исследование носит концептуально-методический и сравнительный характер. Материалом анализа послужили CEFR и CEFR Companion Volume, включая дескрипторы рецепции, продукции, интеракции и медиации [1, 2], рекомендации Совета Европы по соотношению языковых экзаменов с CEFR [3], современная теория валидности результатов тестирования [4], принципы обеспечения качества ALTE [5], спецификации и исследовательские материалы Cambridge C2 Proficiency [6-8], официальные материалы Goethe-Zertifikat C2 [9], действующая структура сертификационного КАЗТЕСТ [10], публикации по продуктивным видам речевой деятельности в КАЗТЕСТ [11], практические

рекомендации ГУ «Кыргызтест» по введению C2 в КАЗТЕСТ [12], а также публикации о развитии национально-ориентированного языкового тестирования в Кыргызстане [13].

Методологически работа опирается на аргументно-ориентированное понимание валидности: объектом доказательства является не тест как таковой, а обоснованность интерпретаций и решений, принимаемых на основе тестового балла [4]. Поэтому предлагаемые структуры, продолжительность, число заданий и веса компонентов рассматриваются как проектные гипотезы. Они требуют эмпирического подтверждения на данных пилотирования.

Применялись четыре группы методов. Во-первых, дескрипторный анализ для выделения наблюдаемых признаков C2 и их разграничения с C1. Во-вторых, сравнительный анализ архитектур экзаменов с акцентом на способы распределения лексико-грамматического контроля. В-третьих, конструктивное проектирование, при котором для каждого блока определялись измерительная цель, тип речевой деятельности, характер стимульного материала и возможные форматы заданий. В-четвертых, экспертно-аналитический синтез рекомендаций для национальной специфики кыргызского и казахского языков.

Принципиальным методическим решением является отказ от трактовки C2 как «уровня носителя языка». CEFR описывает C2 через степень легкости, точности, гибкости и эффективности языкового поведения, а не через происхождение пользователя языка. Для национальных систем это особенно важно: экзамен должен измерять доступную обучению и наблюдаемую языковую компетенцию, а не социально неопределенный образ «идеального носителя».

Результаты и обсуждение

1. Общее концептуальное ядро уровня C2

Сопоставление дескрипторов CEFR показывает, что переход от C1 к C2 следует трактовать прежде всего как качественное изменение характера языковой деятельности. Усложнение не должно сводиться к увеличению длины текста, скорости аудиозаписи, числа терминов или доли редкой лексики. Такие параметры могут повышать поверхностную трудность задания, но не гарантируют измерения более высокого уровня компетенции.

Для C2 определяющими становятся взаимосвязанные характеристики: практически полное понимание сложного дискурса; способность извлекать не только эксплицитную, но и имплицитную информацию; различение тонких смысловых и оценочных оттенков; широкий и точный выбор языковых средств; устойчивый контроль при высокой когнитивной нагрузке; гибкое управление регистром и стилем; интеграция информации из нескольких источников; развернутая, сбалансированная и нюансированная аргументация; способность переформулировать позицию и эффективно управлять интеракцией [1, 2].

Эти характеристики образуют единое ядро и не зависят от того, распределены ли они по четырем или пяти блокам. Следовательно, число субтестов является архитектурным, а не уровневым свойством экзамена.

Таблица 1. Универсальный профиль C2 для национальных систем «Кыргызтест» и КАЗТЕСТ

Измеряемая область	Характеристика C2	Операциональный признак
Понимание	Практически полное и устойчивое	Имплицитные смыслы, позиция, тон, риторическая функция, структура аргументации
Языковой диапазон	Очень широкий и гибкий	Точная номинация, коллокации, перефразирование, синонимический и стилистический выбор

Измеряемая область	Характеристика C2	Операциональный признак
Языковой контроль	Очень высокий	Стабильная точность в сложной задаче и при высокой скорости обработки информации
Дискурс	Связный, гибкий, многослойный	Интеграция источников, ограничение утверждений, причинно-следственные и уступительные связи
Регистр и стиль	Свободная адаптация	Выбор формы с учетом жанра, адресата, статуса и коммуникативной цели
Интеракция	Эффективное управление	Развитие темы, реакция на возражения, уточнение, переговорные и медиативные стратегии
Медиация	Преобразование сложной информации	Обобщение, переупаковка, сопоставление и передача содержания для нового адресата

2. Международный опыт четырехкомпонентной архитектуры

Наиболее показательный пример архитектурной реформы дает Cambridge C2 Proficiency. До 2013 г. экзамен включал пять отдельных компонентов, в том числе Reading и Use of English. В процессе пересмотра экзамена эти два компонента были объединены. Официальные исследовательские материалы Cambridge указывают, что решение было связано с пересечением измерительных фокусов и стремлением сделать экзамен короче и эффективнее без уменьшения охвата ключевых умений [7]. В результате современный C2 Proficiency содержит четыре экзаменационных компонента, но итоговая оценка по-прежнему учитывает Reading, Writing, Listening, Speaking и Use of English как пять равновзвешенных областей результата [6]. Этот пример особенно важен методологически: уменьшение числа экзаменационных блоков не обязательно означает уменьшение числа измеряемых конструктов.

Для КАЗТЕСТ целесообразно применять принцип распределенного измерения языковой формы, при котором лексико-грамматическая компетенция оценивается не как самостоятельный экзаменационный блок, а как компонент выполнения различных видов речевой деятельности. При этом целенаправленные лексико-грамматические задания могут быть интегрированы в блок «Чтение», а проявление языковой точности учитываться при оценивании продуктивных видов речевой деятельности - письма и говорения. Такой подход позволяет сохранить четырехкомпонентную архитектуру КАЗТЕСТ и одновременно обеспечить комплексную оценку языковой компетенции без увеличения количества экзаменационных блоков.

Goethe-Zertifikat C2 использует четыре самостоятельных модуля – чтение, аудирование, письмо и говорение [9]. Важным для национальных систем является не только число модулей, но и характер задач: чтение ориентировано на абстрактные и языково сложные тексты с пониманием имплицитных значений, говорение включает сложное индивидуальное высказывание и дискуссионное взаимодействие, а письмо требует структурной и стилиевой адекватности. Это подтверждает, что четыре макрокомпонента могут охватывать C2 при достаточно глубокой внутренней спецификации.

3. Пятикомпонентная модель C2 для «Кыргызтест»

Пятикомпонентная архитектура «Кыргызтест» может быть сохранена как институционально и методически преемственная модель. При этом C2 требует не добавления более редких языковых единиц, а переопределения измерительного центра каждого субтеста. Самостоятельность «Лексики. Грамматики» оправдана только в том случае, если этот субтест дает уникальную информацию о точности и гибкости языкового контроля и не дублирует содержание чтения и письма.

В субтесте «Лексика. Грамматика» приоритет должны получить контекстуальный выбор, коллокационная точность, вариативность словообразования и грамматических

конструкций, дискурсивные маркеры, модальность, степень уверенности, уступка, ограничение и прагматически мотивированный выбор формы. Задания на редкие исключения, архаизмы и изолированную терминологию не должны становиться основным источником трудности.

«Чтение» на С2 должно быть преимущественно инференциальным и аналитическим. Кандидат реконструирует позицию автора, определяет функцию фрагмента, выявляет логическую организацию и скрытые предпосылки, сопоставляет несколько текстов и интегрирует информацию. «Аудирование» должно оценивать отношение говорящего, неявные связи, изменение позиции, иронию, оговорки и взаимодействие нескольких участников, а не только извлечение фактов.

В «Письме» центральным становится создание нового содержания на основе обработки нескольких источников: обобщение, сопоставление, оценивание аргументов и построение самостоятельного текста с ясными ограничениями и альтернативными трактовками. В «Говорении» необходимо сочетать подготовленное развернутое высказывание с интерактивной дискуссией, где оцениваются развитие позиции, реакция на контраргументы, переформулирование и управление коммуникативным ходом.

4. Четырехкомпонентная модель С2 для КАЗТЕСТ

Сертификационный КАЗТЕСТ в настоящее время имеет четыре блока: аудирование, чтение, письмо и говорение [10]. Практические рекомендации по введению С2, подготовленные ГУ «Кыргызтест», исходят из сохранения этой архитектуры и предлагают качественно пересмотреть содержание каждого блока [12]. Такой подход представляется методически обоснованным, поскольку изменение числа блоков само по себе не является условием соответствия С2.

Ключевой вопрос четырехкомпонентной модели – место лексико-грамматической компетенции. На С2 ее нецелесообразно понимать как набор правил, проверяемых исключительно дискретными заданиями. Более продуктивна распределенная модель, в которой языковой контроль измеряется одновременно в рецептивных и продуктивных действиях. В чтении это может быть понимание функции грамматического выбора, референции, связности, модальности, коллокаций и семантических нюансов в контексте. В письме и говорении – точность, диапазон, уместность, вариативность и способность переформулировать мысль. В аудировании – распознавание прагматической функции интонации, дискурсивных маркеров и грамматически выраженной позиции говорящего.

Таким образом, отсутствие отдельного блока «Лексика. Грамматика» не означает исключения этого конструкта из КАЗТЕСТ. Напротив, на уровне С2 он может измеряться экологически более валидно – через функционирование языковой формы внутри сложной коммуникации. Международный опыт Cambridge показывает, что объединение языкового употребления с чтением может уменьшить избыточность, сохранив самостоятельную интерпретацию показателя Use of English [6, 7]. Для КАЗТЕСТ аналогичный принцип Следует учитывать, что в Cambridge Use of English сохраняется как самостоятельная область результата, тогда как в сертификационном КАЗТЕСТ отдельный лексико-грамматический показатель не выделяется. Поэтому данный пример рассматривается не как структурная модель для прямого заимствования, а как методический ориентир в части интеграции языковой формы в оценивание речевой деятельности. Для КАЗТЕСТ этот принцип может быть реализован с учетом специфики казахского языка – посредством включения лексико-грамматических заданий в текстовые контексты блока «Чтение» и учета языковой точности при оценивании продуктивных видов речевой деятельности. Такой подход позволяет сохранить четырехкомпонентную структуру КАЗТЕСТ без введения отдельного лексико-грамматического блока и обеспечить комплексную оценку языковой компетенции.

При разработке С2 для КАЗТЕСТ особое значение имеет национальное Reference Level Description – детальное описание того, как универсальные дескрипторы С2 реализуются средствами современного казахского языка. Оно должно охватывать лексико-семантическую

точность, устойчивые сочетания, способы выражения модальности и оценки, прагматические маркеры, нормы официально-делового, академического, профессионального и публичного дискурса. Такое описание может стать содержательной основой спецификации и банка заданий.

Таблица 2 - Сопоставительная архитектура С2 в «Кыргызтест» и КАЗТЕСТ

Область	«Кыргызтест»: 5 субтестов	КАЗТЕСТ: 4 блока	Общее требование С2
Лексика и грамматика	Самостоятельный субтест	Распределенный конструкт, интегрированный в оценивание чтения, письма и говорения.	Точность, диапазон, вариативность, контекстуальная уместность
Чтение	Самостоятельный субтест	Самостоятельный блок	Инференция, имплицитность, позиция автора, сопоставление источников
Аудирование	Самостоятельный субтест	Самостоятельный блок	Позиция говорящего, прагматика, естественный темп, сложная полилогическая речь
Письмо	Самостоятельный субтест	Самостоятельный блок	Интеграция источников, аргументация, регистр, языковой контроль
Говорение	Самостоятельный субтест	Самостоятельный блок	Монолог + интеракция, реакция на возражения, управление дискуссией
Медиация	Интегрируется прежде всего в чтение, письмо и говорение	Интегрируется в чтение, письмо и говорение	Преобразование и передача сложной информации

5. Проектные конструкты четырех блоков КАЗТЕСТ С2

Блок «Аудирование». Основной фокус – глубина обработки устного дискурса. Стимулы должны включать лекции, экспертные интервью, аналитические программы, дискуссии, публичные выступления и естественную полилогическую речь. Задания должны различать фактуальное понимание, инференцию, отношение говорящего, степень уверенности, скрытую оценку, иронию, самокоррекцию и логику аргументации. Повышение сложности не следует связывать главным образом с ускорением записи: естественный темп важнее искусственно завышенного темпа.

Блок «Чтение». На уровне С2 целесообразно использовать жанрово неоднородный корпус материалов: тексты аналитического и эссеистического характера, профессиональные и научно-популярные публикации, критико-обзорные и публицистические материалы, в том числе представляющие несовпадающие позиции. Наряду с вопросами на содержание нужны задания на функцию фрагмента, скрытую предпосылку, позицию автора, оценочный оттенок, организацию аргументации и сопоставление нескольких источников. Именно здесь может быть наиболее органично встроена часть лексико-грамматического контроля: выбор точного слова или формы в контексте, интерпретация модальности, референции, связности, коллокаций и смысловых различий.

Блок «Письмо». На С2 письменное задание должно требовать самостоятельной интеллектуальной обработки материала. Перспективной является модель reading-to-write: кандидат получает два или несколько источников, кратко реконструирует их позиции, сопоставляет аргументы и создает новый текст для заданной аудитории и жанра. Критерии оценивания должны включать выполнение коммуникативной задачи, содержание и аргументацию, организацию, лексическую точность и диапазон, грамматический контроль, стилевую уместность, регистровую гибкость и нюансированность позиции. Публикации КАЗТЕСТ по письму и говорению уже подчеркивают необходимость комплексного учета

коммуникативных и языковых параметров [11], что создает методическую основу для повышения требований на С2.

Блок «Говорение». С2 требует выхода за пределы подготовленного монолога. Оптимальна комбинация развернутого индивидуального высказывания и последующей интеракции: уточняющих вопросов, возражений, необходимости защитить, ограничить или переформулировать позицию. Оценивание должно фиксировать не только беглость и правильность, но и способность поддерживать сложную линию аргументации, адаптировать регистр, корректно включаться в дискуссию, переговариваться о значении и управлять ходом коммуникации.

Таблица 3 - Проектная спецификационная рамка КАЗТЕСТ С2

Блок	Ключевой конструкт	Пример задания	Лексико-грамматический контроль	Проектный диапазон времени
Аудирование	Глубокая обработка сложного устного дискурса	Сопоставление позиций нескольких говорящих; вывод об отношении и степени уверенности	Понимание дискурсивных маркеров, модальности и прагматической функции формы	35–50 мин
Чтение	Инференция, структура дискурса, интеграция источников	Сопоставление 2–3 текстов; восстановление логики аргументации; контекстуальный выбор	Встроенные задания на коллокации, связность, словообразование, грамматико-семантические различия	50–70 мин
Письмо	Синтез, аргументация, жанровая и регистровая гибкость	Аналитический текст на основе нескольких источников для заданной аудитории	Аналитический критерий: диапазон, точность, вариативность и уместность формы	60–80 мин
Говорение	Развернутая позиция + интеракция	Краткая подготовка, выступление, вопросы и дискуссия/контраргументация	Аналитический критерий: точность, гибкость перефразирования, стиль и регистр	18–25 мин
<i>Примечание.</i> Временные параметры имеют проектный характер и не должны рассматриваться как нормативные до проведения апробации, анализа нагрузки, надежности и влияния времени на выполнение заданий. Для сравнения: действующее сертификационное тестирование КАЗТЕСТ длится 2 ч 7 мин (аудирование – 20, чтение – 50, письмо – 50, говорение – 7 мин), тогда как суммарное проектное время С2 составляет 163–225 мин.				

6. Разграничение С1 и С2

Практическая сложность разработки высшего уровня заключается в том, что многие языковые средства уже присутствуют на С1. Поэтому спецификация С2 должна фиксировать не новый перечень грамматических тем, а новое качество выполнения. Например, на С1 кандидат способен понимать сложный текст и аргументированно выражать позицию; на С2 от него требуется более устойчивое понимание имплицитности, способность быстро интегрировать несколько перспектив, замечать риторическую функцию формы и строить высказывание, в котором точность сохраняется при высокой концептуальной сложности.

Для чтения разграничивающими заданиями являются вопросы, требующие не воспроизведения информации, а реконструкции отношения автора, скрытого основания аргумента и функции языкового выбора. Для аудирования – интерпретация смены позиции, скрытой оценки и взаимодействия участников. Для письма – способность синтезировать источники и формулировать нюансированную позицию с оговорками и ограничениями. Для

говорения – устойчивость интеракции при неожиданных возражениях и необходимость гибко переформулировать сложную мысль.

Особое значение имеет пограничная зона C1/C2. Она должна быть представлена в корпусе эталонных работ и использоваться при обучении экспертов. Эксперт, уверенно оценивающий уровни A1-C1, не готов автоматически оценивать C2, поскольку на высшем уровне различия часто проявляются в тонких признаках: естественности ограничений, точности регистрового выбора, логике интеграции источников и качестве дискурсивного управления.

7. Валидизация, пилотирование и установление стандарта

Предлагаемые модели должны пройти единый цикл доказательной валидации. Сначала необходимо утвердить описание конструкта C2 и спецификации блоков, затем разработать экспериментальный банк заданий и провести экспертное рецензирование. На этом этапе полезны когнитивные интервью с участниками, позволяющие проверить, действительно ли задание вызывает предполагаемые мыслительные и речевые процессы, а не побочные стратегии угадывания.

Пилотирование целесообразно проводить как минимум на трех группах: устойчивые пользователи C1; предполагаемые пользователи C2; высоко компетентные профессиональные пользователи языка. Сопоставление этих групп позволит оценить, отделяют ли задания C2 от сильного C1. Для объективно оцениваемых заданий необходим анализ трудности и дифференцирующей способности, функционирования дистракторов и возможного дифференциального функционирования заданий. Для письма и говорения – межэкспертная надежность, согласованность критериев и устойчивость баллов при повторном оценивании.

Порог C2 не должен устанавливаться только административным решением или прямым переносом процента правильных ответов с более низких уровней. Требуется стандартизированная процедура *standard setting* с участием подготовленных экспертов, после которой необходимо проверить соответствие классификации содержательному описанию C2 и внешним данным. Особенно важно документировать аргумент валидности: почему именно такой набор заданий, критериев и порогов позволяет делать вывод о владении языком на уровне C2.

Для КАЗТЕСТ дополнительным объектом исследования должна стать распределенная модель лексико-грамматического контроля. Психометрический анализ должен определить, обеспечивает ли лексико-грамматический аналитический критерий и встроенные задания достаточную самостоятельную измерительную информацию, позволяющую надежно оценивать языковую точность без введения отдельного лексико-грамматического блока. При выявлении существенных ограничений или недостаточной измерительной эффективности спецификация может быть пересмотрена; при этом соответствующие решения должны основываться на эмпирических данных, а не на предположении о необходимости отдельного лексико-грамматического субтеста для уровня C2.

В КАЗТЕСТ пороговые требования устанавливаются по каждому из четырех блоков сертификационного тестирования - аудированию, чтению, письму и говорению. В связи с этим при разработке модели уровня C2 целесообразно сохранить данный принцип блочного оценивания. При этом лексико-грамматический контроль, интегрированный в блок «Чтение», а также учитываемый при оценивании письма и говорения, не предусматривает установления отдельного порогового значения, поскольку не выделяется в самостоятельный экзаменационный блок.

Для обеих систем необходимо сформировать корпус эталонных письменных и устных ответов C1, пограничного C1/C2 и устойчивого C2, провести специальную калибровку экспертов и на этапе апробации использовать двойное независимое оценивание продуктивных видов речевой деятельности. После запуска следует вести ежегодный

мониторинг параметров заданий, стабильности порогов, межэкспертного согласия, групповых различий и последствий использования сертификата.

Таблица 4 - Рекомендуемая этапность разработки и внедрения C2

Этап	Основная задача	Ключевой результат	Доказательство качества
1. Концептуализация	Описание C2 и разграничение C1/C2	Национальные дескрипторы и матрица конструктора	Экспертное согласование
2. Спецификация	Определение типов заданий, критериев и покрытия	Спецификации 5 или 4 компонентов	Контентная валидность
3. Экспериментальный банк	Создание и рецензирование материалов	Пул заданий и шкал	Экспертные заключения, когнитивные интервью
4. Пилотирование	Получение эмпирических данных	Параметры заданий и работ кандидатов	Психометрический анализ
5. Стандартизация	Обучение экспертов и установление порога	Эталонные работы и cut score	Межэкспертная надежность, standard setting
6. Ограниченное внедрение	Проверка операционных процедур	Стабильная форма экзамена	Мониторинг администрирования и последствий
7. Полномасштабное внедрение	Регулярная сертификация	Операционная система C2	Ежегодный анализ и пересмотр

8. Национальная специфика кыргызского и казахского C2

Универсальные дескрипторы CEFR должны быть конкретизированы применительно к функциональной системе каждого языка. Для кыргызского и казахского языков важно учитывать не только формально-грамматические особенности, но и нормы современного официально-делового, академического, профессионального, медийного и публичного дискурса. Следовательно, при разработке заданий необходимо опираться на актуальные корпуса и реальные жанры, а не создавать искусственно усложненные тексты, насыщенные мало употребительными формами.

Для C2 особенно значима социолингвистическая и прагматическая точность: способность выбирать форму обращения, степень категоричности, способы смягчения или усиления утверждения, корректно маркировать источник информации, позицию и оценку. В агглютинативных языках дополнительное внимание требуется к функциональной нагрузке морфологических средств, цепочкам словообразования, синтаксической вариативности и связности развернутого текста. Однако эти особенности должны вводиться в спецификацию как функциональные признаки, а не как перечень «сложных грамматических тем».

Национальные Reference Level Descriptions для C2 могут стать связующим звеном между CEFR и тестовой спецификацией. Они позволят зафиксировать типичные лексические, грамматические, дискурсивные и прагматические реализации высокого уровня именно в кыргызском и казахском языках и тем самым повысить прозрачность разработки заданий и обучения экспертов.

9. Ограничения исследования

Предлагаемые модели не являются готовыми валидированными экзаменами. На момент исследования отсутствуют эмпирические данные по операционному уровню C2 в обеих системах, поэтому числовые параметры, типы заданий и распределение времени имеют проектный характер. Дополнительного исследования требует влияние цифрового формата: скорость набора текста, особенности интерфейса, качество воспроизведения

аудиоматериалов и уровень цифровой грамотности могут влиять на результаты независимо от уровня языковой компетенции.

Для КАЗТЕСТ предусматривается распределенная модель лексико-грамматического контроля. Лексико-грамматические задания интегрируются в блок «Чтение», при этом языковая точность также учитывается при оценивании письма и говорения. Такой подход позволяет сохранить четырехкомпонентную структуру КАЗТЕСТ и осуществлять комплексную оценку языковой компетенции без выделения отдельного лексико-грамматического блока.

В отличие от этого, сохранение отдельного субтеста «Лексика. Грамматика» в «Кыргызтест» требует подтверждения его самостоятельного измерительного вклада в итоговую классификацию.

Заключение

Разработка уровня C2 для «Кыргызтест» и КАЗТЕСТ должна начинаться не с создания более трудных заданий, а с ясного описания качественно иного уровня языковой деятельности. C2 характеризуется глубиной понимания, интерпретацией имплицитных смыслов, точностью и гибкостью языкового выбора, устойчивым регистровым контролем, интеграцией информации, нюансированной аргументацией, медиацией и эффективным управлением взаимодействием.

Исследование показывает, что эти характеристики могут быть операционализированы в разных архитектурах. Для «Кыргызтест» методически возможно сохранение пяти субтестов, включая самостоятельную «Лексику. Грамматику», если каждый компонент имеет четко определенный и не дублирующий конструкт. Для КАЗТЕСТ обосновано сохранение действующей для сертификационного КАЗТЕСТ целесообразно сохранить действующую четырехкомпонентную структуру - аудирование, чтение, письмо и говорение - с интеграцией лексико-грамматического контроля в структуру теста без выделения отдельного лексико-грамматического блока.

Опыт Cambridge C2 Proficiency, перешедшего от пяти компонентов к четырем за счет объединения Reading и Use of English, подтверждает, что архитектурное укрупнение возможно без потери измерительного содержания при условии осознанной перестройки спецификации. Четырех модульная модель Goethe-Zertifikat C2 дополнительно демонстрирует возможность полноценного оценивания высшего уровня через четыре макро вида речевой деятельности.

Следующим этапом для обеих национальных систем должна стать разработка детальных спецификаций C2, экспериментальных банков заданий, национальных описаний уровня, корпусов эталонных ответов и программ подготовки экспертов. До массовой сертификации необходимы многоэтапное пилотирование, психометрический анализ, проверка надежности продуктивных блоков, процедура установления порогов и документированное доказательство валидности. Именно такая последовательность позволяет соединить национальную языковую специфику Кыргызстана и Казахстана с международно признанными принципами языкового оценивания.

Список литературы

1. Council of Europe. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press. - 2001.
2. Council of Europe. Common European Framework of Reference for Languages: Learning, Teaching, Assessment - Companion Volume. Strasbourg: Council of Europe. - 2020.
3. Council of Europe. Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Strasbourg: Council of Europe. - 2009.

4. Kane M.T. Validating the Interpretations and Uses of Test Scores. Journal of Educational Measurement. 2013. - 50(1). - P. 1-73. - <https://doi.org/10.1111/jedm.12000>
5. Association of Language Testers in Europe (ALTE). Principles of Good Practice. 3rd ed. Cambridge: ALTE. - 2020.
6. Cambridge University Press & Assessment. C2 Proficiency: Handbook for Teachers. Cambridge. - 2025. - <https://www.cambridgeenglish.org/Images/168194-c2-proficiency-teachers-handbook.pdf>
7. Cambridge ESOL. The revised Cambridge English: Proficiency examination. Research Notes. 2012. - (49). - P. 3-17. Available at: <https://www.cambridgeenglish.org/images/23166-research-notes-49.pdf>
8. Cambridge University Press & Assessment. C2 Proficiency exam format. - <https://www.cambridgeenglish.org/exams-and-tests/qualifications/proficiency/format/>
9. Goethe-Institut. Goethe-Zertifikat C2: Großes Deutsches Sprachdiplom. Further information / Exam modules. - <https://www.goethe.de/ins/de/en/m/prf/prf/gzc2/inf.html>
10. РГП на ПХВ «Национальный центр тестирования» МНВО Республики Казахстан. Сертификационное тестирование на оценку уровня владения казахским языком КАЗТЕСТ: структура тестирования, Астана. - 2026. - https://new.testcenter.kz/?lang=ru&page_id=21316
11. Тоғызбай Н.Б., Ерсұлтанова Г.С., Жумаханова Н.Н. Особенности продуктивной речевой деятельности в системе КАЗТЕСТ. Вестник Евразийского гуманитарного института. - 2026. - №1. - <https://doi.org/10.55808/1999-4214.2026-1.33>
12. Государственное учреждение «Кыргызтест». Практические рекомендации по внедрению уровня C2 в систему оценки владения казахским языком КАЗТЕСТ, Бишкек. - 2026.
13. Степаненко В.А., Нахабина М.М., Топоева-Ставинская Б.Б. Тест по русскому языку как официальному в Кыргызстане: становление и развитие национальной системы тестирования. Русский тест: теория и практика. - 2024. - 10(1). - С. 3-12. - <https://doi.org/10.22363/3034-2090-2024-10-1-3-12>

Б.Б. Топоева-Ставинская, Г.С. Ерсұлтанова

КЫРГЫЗТЕСТ ЖӘНЕ ҚАЗТЕСТ ҰЛТТЫҚ ТІЛДІК ТЕСТІЛЕУ ЖҮЙЕЛЕРІНДЕ C2 ДЕҢГЕЙІН ӨЗІРЛЕУДІҢ ТҰЖЫРЫМДАМАЛЫҚ-ӘДІСНАМАЛЫҚ НЕГІЗДЕРІ: БЕС ЖӘНЕ ТӨРТ КОМПОНЕНТТІ ҚҰРЫЛЫМДАРДЫҢ САЛЫСТЫРМАЛЫ МОДЕЛІ

Зерттеудің өзектілігі Қырғызстан мен Қазақстанның ұлттық тілдік бағалау шкалаларын халықаралық тәсілдермен салыстырмалылықты және тілдердің ұлттық ерекшеліктерін сақтай отырып, ең жоғары C2 деңгейіне дейін кеңейту қажеттілігімен айқындалады. Жұмыстың мақсаты – «Кыргызтест» және КАЗТЕСТ жүйелері үшін C2 деңгейінің ортақ тұжырымдамалық негізін айқындау және екі операциялық модельді қарастыру: «Кыргызтест» үшін бес компонентті және КАЗТЕСТ үшін төрт компонентті модель. Әдіснама CEFR және CEFR Companion Volume дескрипторларын талдауды, Cambridge C2 Proficiency және Goethe-Zertifikat C2 қолданыстағы форматтарын салыстыруды, сондай-ақ Cambridge Proficiency емтиханы реформасын талдауды қамтиды. Аталған реформа нәтижесінде бес компонентті емтихан Reading және Use of English бөлімдерін біріктіру арқылы төрт компонентті форматқа өзгертілді. Сонымен қатар екі ұлттық жүйенің қолданыстағы құрылымдары негізінде конструктивті жобалау жүргізілді. Зерттеу нәтижелері C2 деңгейі C1 деңгейінен материалдың сандық тұрғыдан күрделенуімен емес, тілдік әрекеттің сапасымен ерекшеленуі тиіс екенін көрсетеді. Бұл сапа түсінудің тереңдігімен, имплицитті мағыналарды интерпретациялау қабілетімен, тілдік құралдарды дәл таңдаумен, регистрлік икемділікпен, әртүрлі дереккөздерді ықпалдастырумен, нюансты аргументациямен және өзара әрекеттесуді басқарумен сипатталады. КАЗТЕСТ-тің сертификаттық тестілеу форматы төрт блоктан: тыңдалым, оқылым, жазылым және айтылымнан тұрады. Лексика-грамматикалық құзыреттілікті бағалау «Оқылым» блогына енгізілген кіріктірілген тапсырмалар арқылы жүзеге асырылады. «Кыргызтест» жүйесінде бес компонентті модельдің құрамдас бөлігі ретінде «Лексика. Грамматика» дербес субтесті сақталады.

Түйін сөздер: CEFR, C2 деңгейі, тілдік тестілеу, «Кыргызтест», КАЗТЕСТ, Ұлттық бағалау жүйесі, валидтілік, тест спецификациясы, психометриялық валидация, төрт компонентті модель.

B.B. Topoeva-Stavinskaya, G.S. Yersultanova

CONCEPTUAL AND METHODOLOGICAL FOUNDATIONS FOR DEVELOPING C2 IN THE KYRGYZTEST AND QAZTEST NATIONAL LANGUAGE ASSESSMENT SYSTEMS: A COMPARATIVE MODEL OF FIVE- AND FOUR-COMPONENT ARCHITECTURES

The relevance of the study is determined by the need to extend the national language assessment scales of Kyrgyzstan and Kazakhstan to the highest proficiency level, C2, while maintaining comparability with international approaches and preserving the national and linguistic specificities of the respective languages. The aim of the study is to substantiate a common conceptual framework for C2 within the Kyrgyztest and QAZTEST systems and to examine two operational models: a five-component model for Kyrgyztest and a four-component model for QAZTEST. The methodology includes a descriptor-based analysis of the CEFR and the CEFR Companion Volume, a comparison of the current formats of Cambridge C2 Proficiency and Goethe-Zertifikat C2, an analysis of the Cambridge Proficiency reform, whereby the former five-component examination was transformed into a four-component format through the integration of Reading and Use of English, as well as construct-oriented test design based on the existing architectures of the two national assessment systems. The study demonstrates that C2 should differ from C1 not through a quantitative increase in task difficulty, but through qualitative characteristics of language performance, including depth of comprehension, interpretation of implicit meanings, precision in the selection of linguistic resources, register flexibility, integration of information from multiple sources, nuanced argumentation, and management of interaction. The QAZTEST certification testing format consists of four components: Listening, Reading, Writing, and Speaking. Lexical and grammatical competence is assessed through integrated tasks incorporated into the Reading component. Kyrgyztest, meanwhile, retains an independent “Vocabulary and Grammar” subtest as part of its five-component model. It is concluded that both architectures can support a valid interpretation of C2, provided that they are supported by detailed test specifications, multi-stage piloting, psychometric validation, standardization of assessment procedures, and formal standard-setting procedures for establishing cut scores.

Keywords: CEFR, C2, language assessment, Kyrgyztest, QAZTEST, National assessment system, validity, test specification, psychometric validation, four-component model.

References

1. Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
2. Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
3. Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A manual*. Council of Europe.
4. Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
5. Association of Language Testers in Europe. (2020). *Principles of good practice* (3rd ed.). ALTE.
6. Cambridge University Press & Assessment. (2025). *C2 Proficiency: Handbook for teachers*. <https://www.cambridgeenglish.org/Images/168194-c2-proficiency-teachers-handbook.pdf>
7. Cambridge ESOL. (2012). The revised Cambridge English: Proficiency examination. *Research Notes*, 49, 3–17. <https://www.cambridgeenglish.org/images/23166-research-notes-49.pdf>
8. Cambridge University Press & Assessment. (2026). *C2 Proficiency exam format*. Retrieved September 30, 2026, from <https://www.cambridgeenglish.org/exams-and-tests/qualifications/proficiency/format/>
9. Goethe-Institut. (2026). *Goethe-Zertifikat C2: Großes Deutsches Sprachdiplom: Further information*. Retrieved September 30, 2026, from <https://www.goethe.de/ins/de/en/m/prf/prf/gzc2/inf.html>
10. MNVO Respubliki Kazakhstan. (2026). *Respublikanskoe gosudarstvennoe predpriyatie na prave khozyaistvennogo vedeniya “Natsionalnyi tsentr testirovaniya” Sertifikatsionnoe testirovanie na otsenku urovnya vladeniya kazakhskim yazykom KAZTEST: struktura testirovaniya* [Certification testing for

assessing Kazakh language proficiency in KAZTEST: Test structure]. National Center for Testing. https://new.testcenter.kz/?lang=ru&page_id=21316

11. Togyzbay, N. B., Yersultanova, G. S., & Zhumakhanova, N. N. (2026). Osobennosti produktivnoi rechevoi deyatelnosti v sisteme KAZTEST [Features of productive speech activity in the KAZTEST system]. *Vestnik Evraziiskogo Gumanitarnogo Instituta*, (1), 33. <https://doi.org/10.55808/1999-4214.2026-1.33>

12. Gosudarstvennoe uchrezhdenie “Kyrgyztest”. (2026). *Prakticheskie rekomendatsii po vnedreniyu urovnya C2 v sistemu otsenki vladeniya kazakhskim yazykom KAZTEST* [Practical recommendations for introducing C2 into the KAZTEST system for assessing Kazakh language proficiency] [Unpublished manuscript].

13. Stepanenko, V. A., Nakhabina, M. M., & Topoeva-Stavinskaya, B. B. (2024). Test po russkomu yazyku kak ofitsialnomu v Kyrgyzstane: stanovlenie i razvitie natsionalnoi sistemy testirovaniya [Test of Russian as an official language in Kyrgyzstan: Formation and development of the national testing system]. *Russian Language: Research, Testing and Practice*, 10(1), 3–12. <https://doi.org/10.22363/3034-2090-2024-10-1-3-12>

Сведения об авторах:

Топоева-Ставинская Батма Болоталиевна (автор-корреспондент) – заслуженный работник образования Кыргызской Республики, директор Государственного учреждения «Кыргызтест», основатель уровневой системы «Кыргызтест», Бишкек, Кыргызская Республика, e-mail: info@kyrgyztest.gov.kg

Ерсултанова Гульнур Серикбаевна – PhD-докторант, руководитель Управления по развитию системы КАЗТЕСТ, РГП на ПХВ «Национальный центр тестирования» Министерства науки и высшего образования Республики Казахстан, Астана, Казахстан, e-mail: ersultanovagulnur@gmail.com

Авторлар туралы мәлімет:

Топоева-Ставинская Батма Болоталиевна (автор-корреспондент) – Кыргыз Республикасынын билим беру саласына еңбек сіңірген қызметкері, «Қырғызтест» мемлекеттік мекемесінің директоры, «Қырғызтест» деңгейлік жүйесінің негізін қалаушы, Бішкек, Қырғыз Республикасы, e-mail: info@kyrgyztest.gov.kg

Ерсултанова Гульнур Серикбаевна – PhD докторант, ҚАЗТЕСТ жүйесін дамыту жөніндегі басқарма басшысы, Қазақстан Республикасы Ғылым және жоғары білім министрлігінің «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail: ersultanovagulnur@gmail.com

Information about the authors:

Topoeva-Stavinskaya Batma Bolotalievna (corresponding author) – Honored Worker of Education of the Kyrgyz Republic, Director of the State Institution “Kyrgyztest”, founder of the Kyrgyztest proficiency-level system, Bishkek, Kyrgyz Republic, e-mail: info@kyrgyztest.gov.kg

Yersultanova Gulnur Serikbayevna – PhD doctoral student, Head of the Department for the Development of the KAZTEST System, National Testing Center of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan, e-mail: ersultanovagulnur@gmail.com

Ш.Б. Алтыбаева^{1*}, А.Н. Қуанышбекова², Ж. Абдулла³, А.Н. Есимбекова⁴

^{1,2,4} РГП на ПХВ «Национальный центр тестирования» МНВО РК, г. Астана, Республика Казахстан

³ КГУ «Специализированный лицей № 82 «Дарын»», г. Астана, Республика Казахстан

*e-mail: shygyla.altymbayeva@gmail.com

¹ORCID 0000-0003-0306-861X, ²ORCID 0009-0000-3780-6866,

³ORCID 0009-0004-9229-6108, ⁴ORCID: 0009-0005-3896-5654

ОЦЕНКА КОГНИТИВНОЙ СЛОЖНОСТИ ЗАДАНИЙ ПО ХИМИИ: ДАННЫЕ ЕНТ КАЗАХСТАНА

В настоящей работе оценивается, насколько две формы ЕНТ по химии сопоставимы не только по содержанию, но и по характеру познавательных действий, которые требуются от абитуриента. Проанализировано 80 заданий двух форм и классифицировали их когнитивную сложность по трехуровневой версии Depth of Knowledge (DOK) Уэбба. Шесть экспертов независимо кодировали все задания. На первом этапе абсолютное совпадение экспертных оценок составило 87%; случаи расхождения мы затем разбирали совместно и определяли итоговую категорию по консенсусу. В форме «Химия-1» к DOK-1, DOK-2 и DOK-3 мы отнесли соответственно 35,0%, 40,0% и 25,0% заданий, а в форме «Химия-2» - 57,5%, 25,0% и 17,5%. Наибольшее расхождение пришлось на DOK-1 и составило 22,5 процентного пункта. Дополнительно мы сопоставили DOK с эмпирической трудностью по модели Раша. Средняя трудность возрастала от DOK-1 к DOK-3, однако диапазоны отдельных уровней перекрывались. На основании этого результаты показывают, что одинаковое количественное распределение заданий по содержательным разделам еще не означает одинаковых когнитивных требований. В рамках данного пилотного анализа DOK-профиль оказался полезной дополнительной характеристикой тестовой формы. Поскольку мы рассмотрели только два варианта, выводы требуют проверки на более широком материале.

Ключевые слова: когнитивная сложность, DOK, ЕНТ, химия, high-stakes testing, валидность, модель Раша, параллельность форм.

Введение

В статье исходным положением является то, что для экзамена с высокими ставками недостаточно подтвердить надежность и соответствие теста содержательной спецификации. Важно также определить, какие действия должен выполнить абитуриент, чтобы получить правильный ответ. Этот аспект связан с представлением измеряемого конструкта и, следовательно, с обоснованностью интерпретации результатов [1, 2]. Если форма в основном проверяет воспроизведение фактов, она дает меньше информации о способности применять знания и рассуждать.

В исследовании разграничиваются когнитивная сложность и эмпирическую трудность задания. Под когнитивной сложностью под когнитивной сложностью понимается характер требуемого действия, тогда как трудность проявляется в результатах конкретной выборки. На *p*-value могут влиять формулировка, знакомость контекста, формат задания, подготовка участников и другие факторы [3, 4]. Поэтому когнитивный анализ когнитивный анализ рассматривается как самостоятельный слой описания теста. Сходную задачу решают исследования, в которых используются таксономия Блума и когнитивно-диагностические модели [5].

Для описания этого слоя мы выбрали модель Depth of Knowledge Уэбба и использовали трехуровневую схему: DOK-1 - воспроизведение, DOK-2 - применение знаний, навыков и понятий, DOK-3 - стратегическое мышление [6, 7]. При классификации учитывался не внешний вид задания и не число шагов решения, а то действие, без которого нельзя получить правильный ответ.

Возможность такого подхода к химическим заданиям показана, в частности, в работах Holme et al. и Raker et al. [8, 9].

В ЕНТ результаты тестирования используются при отборе абитуриентов в организации высшего образования. Поэтому, рассматривая параллельные формы, важно контролировать не только одинаковый охват тем. Даже при формальном соответствии одной спецификации формы могут различаться по тому, как часто они требуют воспроизведения, применения известного способа или самостоятельного выбора стратегии.

На материале химии это различие особенно удобно наблюдать. Часть заданий решается через воспроизведение определения, формулы или правила. В других случаях абитуриенту нужно сопоставить вещества и процессы, прочитать числовые данные, выбрать расчетный путь или объяснить химическое явление. Поэтому в анализе тематическая матрица не заменяет оценку характера познавательной работы.

При формировании аналитической схемы мы опирались на действующую спецификацию ЕНТ по химии. Она включает задания по строению и частицам вещества, закономерностям протекания реакций, энергетике, разделам «Химия и жизнь» и «Химия вокруг нас», а также расчетные и практико-ориентированные задания [10]. Это позволило нам отдельно рассматривать тематическое представление материала и требуемые от тестируемого действия.

При обзоре казахстанских публикаций мы обнаружили, что близкие по смыслу вопросы чаще рассматриваются в контексте учебной практики. Садыков, Кокибасова и Нурмыш, например, разработали разноуровневые задания для 9 класса, ориентированные на естественно-научную грамотность [11]. Авторы отмечают затруднения школьников при расчетах, составлении уравнений реакций и ответах на проблемные вопросы и предлагают различать задания по содержанию, характеру и объему работы.

Работа Шаңытбаевой также оказалась значимой для нашего обзора: в ней познавательные задания рассматриваются как средство формирования химических знаний [12]. Автор связывает их с развитием мышления, соединением теории и практики и элементами исследовательской деятельности. Для рассматриваемой задачи эта публикация важна прежде всего тем, что показывает интерес к характеру познавательного действия в казахстанском химическом образовании, хотя сравнение параллельных экзаменационных форм там не проводилось.

В международной литературе мы нашли примеры более формализованного анализа экзаменационных материалов. Okanlawon и Adeoti распределили 230 заданий пяти выпускных экзаменов West African Senior School Certificate Examination по категориям Revised Bloom's Taxonomy [13]. Такой дизайн позволил авторам количественно описать когнитивный состав реальных экзаменационных заданий, а не только учебных упражнений.

Еще один ориентир для нашей работы - исследование Zorluoğlu, Kizilaslan и Dönmez Yarıcıoğlu [14], где Revised Bloom's Taxonomy применена к программе и учебнику химии 9 класса. Авторы отнесли 61% результатов обучения к understand, 16% - к remember, 5% - к apply, 13% - к analyse, 0% - к evaluate и 5% - к create. Здесь важна сама возможность количественно описывать когнитивную структуру содержания.

Ren et al. анализировали пять форм китайского Гаокао по химии с помощью предметно-специфической модели компетенций [15] и выявили различия между формами по характеру требуемой деятельности. В качестве инструмента выбран - DOK, поскольку наша задача уже: сопоставить глубину требуемых действий в двух формах одного экзамена по единой шкале.

По результатам обзора мы выделили два основных направления: исследования учебных заданий по химии и работы, посвященные когнитивным требованиям реальных экзаменационных материалов. Для ЕНТ по химии нам не удалось обнаружить большого числа исследований, где несколько форм систематически кодировались бы по одной когнитивной шкале. Поэтому анализ сосредоточен на прямом сравнении двух реальных форм.

Практический смысл такого сравнения заключается. Спецификация показывает, какие темы и проектные уровни трудности должны быть представлены, но сама по себе не отвечает на вопрос, какую мыслительную операцию требует конкретное задание. С помощью DOK мы фиксировали эту характеристику отдельно от содержания и предполагаемой трудности.

Материалы и методы исследования

Исследование построено как пилотное сравнительное. Единицей анализа было отдельное задание ЕНТ по химии. Сначала мы кодировали задания двух форм по DOK, затем сопоставляли полученные когнитивные характеристики с содержанием, форматом и психометрическими показателями.

В анализ включены 80 заданий - по 40 из форм «Химия-1» и «Химия-2». По обеим формам были доступны экспертные оценки и данные выполнения заданий. Это позволило не ограничиваться содержательным описанием и проверить, как экспертно присвоенные уровни DOK соотносятся с наблюдаемой трудностью.

Использованы две реальные апробационные формы одного экзаменационного периода. Отбор был целенаправленным: включены варианты с полными эмпирическими данными, соответствовавшие действующей спецификации по содержанию и проектному распределению трудности. Случайную выборку форм мы на этом этапе не формировали.

Такое решение ограничивает область наших выводов. На основании двух форм результаты не переносятся на все варианты ЕНТ по химии; здесь мы прежде всего проверяем рабочую процедуру анализа. Каждая форма содержала задания с одним правильным ответом, контекстные задания, задания с несколькими ответами и задания на установление соответствия. Проектное соотношение категорий А, В и С составляло 50%, 30% и 20%.

В работе использовались сами тестовые задания, действующую спецификацию, контрольные листы разработки и внутренние материалы экспертной оценки Национального центра тестирования.

Операционализация когнитивной сложности

Для кодирования мы адаптировали модель Webb Depth of Knowledge [6, 7]. В рабочей версии оставили три уровня, которых, по нашему мнению, достаточно для различения основных типов действий в рассматриваемых заданиях по химии:

- DOK-1 - воспроизведение. К этому уровню мы относили задания, где правильный ответ можно получить за счет знания факта, определения, формулы, правила или выполнения простой освоенной процедуры.
- DOK-2 - применение. Этот уровень мы присваивали заданиям, в которых тестируемому нужно использовать знание или известный способ в конкретной либо измененной ситуации, интерпретировать, преобразовать или обработать представленную информацию.
- DOK-3 - стратегическое мышление. Сюда мы относили задания, где требуется связать несколько элементов, выбрать способ решения, выстроить содержательное рассуждение или применить стратегию, которая не задается напрямую стандартным алгоритмом.

При присвоении уровня учитывалось на минимально необходимое когнитивное действие, а не на формат задания. Категория А/В/С, объем текста и число расчетных шагов сами по себе уровень не повышали. Например, длинный расчет мы относили к DOK-2, если последовательность действий полностью задавалась известной процедурой.

Поэтому далее мы используем DOK как характеристику глубины действия, заложенного в задании, а не как другое обозначение эмпирической трудности

Экспертное кодирование

В экспертной процедуре участвовали шесть человек - специалисты Национального центра тестирования и обученные эксперты ETS. Каждый эксперт самостоятельно классифицировал все 80 заданий.

Перед началом работы мы предоставили экспертам единые описания трех уровней, адаптированные к химическому содержанию. В инструкции отдельно указали, что сложная формула, большой объем условия, несколько вычислений или проектная категория трудности сами по себе не являются основанием для присвоения более высокого DOK.

Мы предложили следующую последовательность кодирования: сначала эксперт формулировал действие, без которого задание нельзя решить, и только затем выбирал уровень DOK. Воспроизведение факта или правила относили к первому уровню; применение, интерпретацию и преобразование информации - ко второму; выбор стратегии и интеграцию нескольких взаимосвязанных элементов - к третьему.

Особенно внимательно мы рассматривали границу между DOK-2 и DOK-3. Мы не приравнивали многоступенчатое решение к стратегическому мышлению. Если учащийся мог выполнить задание по знакомой последовательности действий, мы оставляли его на уровне DOK-2.

По каждому заданию эксперт фиксировал выбранный уровень и краткое обоснование. Одновременно мы регистрировали содержательный раздел, способ измерения, проектную категорию A/B/C, тип стимульного материала и формат ответа. В дальнейшем эти признаки использовали для сопоставлений.

После независимого кодирования оценки были сведены в одну таблицу. Полное совпадение всех шести оценок на первом этапе наблюдалось в 87% случаев. Остальные задания проводилось совместное обсуждение: возвращались к формулировке задания и описанию требуемого действия, после чего принимали итоговую категорию по консенсусу.

Первоначальные оценки мы использовали для характеристики согласованности экспертной процедуры, а в последующий количественный анализ включали уже единый консенсусный код каждого задания.

Мы понимаем ограничение такой классификации: она описывает требование, заложенное в задании, но не фактический ход мысли каждого участника. Без когнитивных интервью нельзя утверждать, что все тестируемые решали одно и то же задание одинаковым способом. Поэтому DOK мы интерпретируем на уровне дизайна форм.

Анализ распределения когнитивной сложности. Для каждой формы мы подсчитали абсолютное число и процент заданий DOK-1, DOK-2 и DOK-3. Дополнительно объединили DOK-2 и DOK-3, чтобы получить простой описательный показатель того, насколько часто форма требует действий, выходящих за рамки непосредственного воспроизведения.

Разницу между формами мы выражали в процентных пунктах. Порог 10 процентных пунктов использовали только как внутренний ориентир для пилотного сравнения и не трактовали его как установленную границу эквивалентности форм.

Анализ проектной и эмпирической трудности. Проектную трудность определялась по спецификации в виде категорий A, B и C. Наблюдаемую трудность описывали через p-value - долю участников, которые дали правильный ответ.

Сначала мы сравнили распределения p-value внутри категорий A, B и C, а затем сопоставили те же категории с DOK. Принципиально различать эти признаки: A/B/C задает предполагаемую трудность на этапе конструирования, а DOK описывает характер требуемого познавательного действия.

Психометрический анализ. Для дополнительной характеристики исследуемых форм был проведен анализ на основе модели Раша [22] с использованием программного обеспечения Winsteps 5.8.3.0. Выборки абитуриентов, выполнявших формы «Химия-1» (N=121) и «Химия-2» (N=161), формировались независимо в рамках регулярного тестирования. Различие в объеме обусловлено организационными причинами и не связано с содержанием заданий. Для целей пилотного исследования такой разброс допустим. Размеры выборок различались, поэтому психометрические показатели двух форм интерпретировались преимущественно описательно [23, 24]. При данных объемах стандартные ошибки оценок трудности заданий различаются не более чем на 0,02 logit, что не влияет на интерпретацию

показателей соответствия (infit/outfit MNSQ) и не ставит под сомнение сопоставимость психометрических результатов двух форм.

Обеспечение качества и интерпретация результатов. Для контроля качества процедуры мы использовали три элемента: независимое кодирование шестью экспертами, единую инструкцию и обязательное обсуждение несовпадающих оценок. Наиболее внимательно мы контролировали границу DOK-2/DOK-3, поскольку именно здесь сложное содержание или длинный расчет могут привести к завышению уровня.

Психометрические показатели мы рассматриваем как вспомогательный источник информации. Они позволяют увидеть, как задания с разными DOK ведут себя в эмпирических данных, но не доказывают причинного влияния когнитивного уровня на трудность или надежность формы. По той же причине экспертный DOK не следует интерпретировать как прямое наблюдение реальных мыслительных процессов участников.

Результаты и обсуждение

В количественном анализе были использованы три уровня DOK: воспроизведение (DOK-1), применение знаний и навыков (DOK-2) и стратегическое мышление (DOK-3) [6, 7]. Краткие рабочие определения, которыми мы руководствовались, приведены в таблице 1.

Таблица 1 – Уровни DOK, использованные при анализе заданий по химии

Уровень	Когнитивная сложность
DOK 1	Вспомнить и воспроизвести
DOK 2	Применить знание, навык, понятие
DOK 3	Стратегическое мышление

Предварительно мы просмотрели распределение DOK по профильным предметам и увидели, что задания заметно различаются по характеру требуемых действий. В химии была представлена заметная доля DOK-3 (рисунок 1). Далее мы анализируем только две химические формы, поскольку именно по ним располагали полным набором экспертных кодов и эмпирических показателей. В обеих формах встречались задания от простого воспроизведения до задач, где требовалось анализировать связи и выбирать способ решения.

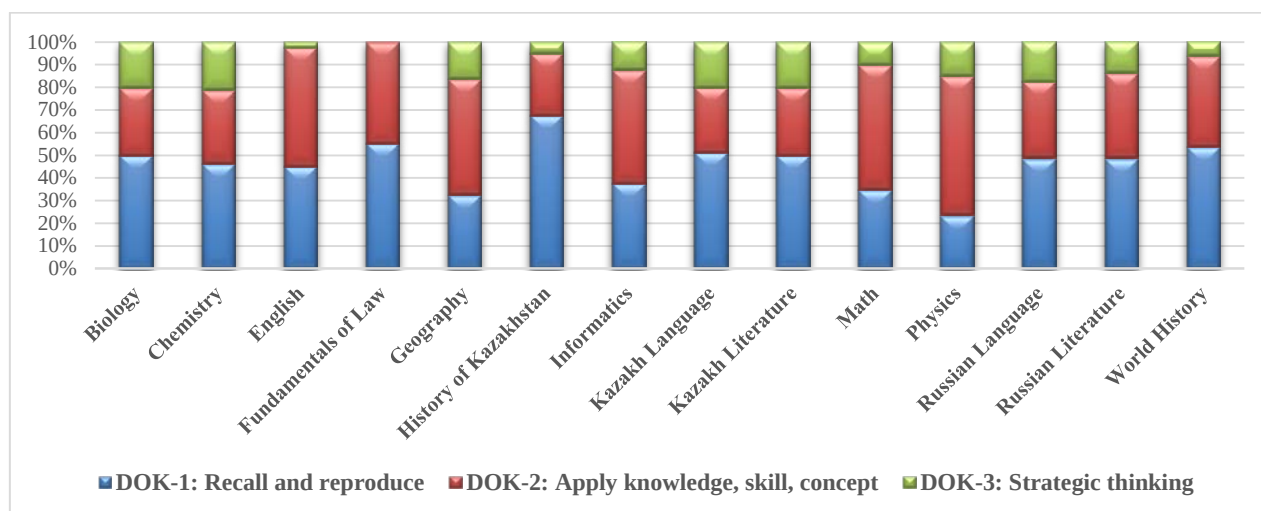


Рисунок 1 – Распределение заданий по уровням когнитивной сложности DOK

Такое сочетание уровней в химии мы считаем закономерным. Часть содержания действительно проверяется через факты, формулы и определения, но значительный блок заданий требует применять закономерности, читать числовую информацию, связывать вещества и процессы и выполнять расчеты. При этом наличие расчета само по себе мы не считали признаком DOK-3: решающим оставался характер рассуждения.

Задача сравнения разные предметы по единой «правильной» пропорции DOK. У химии, физики и биологии различаются предметные задачи и типичные способы решения. Поэтому интерес представляет не абстрактно высокий или низкий уровень когнитивной сложности, а согласованность профиля с целями конкретного теста.

Две формы сопоставлялись по итоговым экспертным кодам всех 40 заданий каждой формы.

Различие видно уже по частотам. В «Химии-1» мы отнесли к DOK-1 14 заданий (35,0%), к DOK-2 - 16(40,0%), к DOK-3 - 10 (25,0%). В «Химии-2» соответствующие значения составили 23 (57,5%), 10 (25,0%) и 7 (17,5%) (таблица 2).

Таблица 2 – Распределение уровней DOK по предмету «Химия» по данным содержательного анализа НЦТ

Профиль (№ формы)	DOK-1, %	DOK-2, %	DOK-3, %
Химия-1	35,0	40,0	25,0
Химия-2	57,5	25,0	17,5

Примечание: Cramér's $V = 0,23$ указывает на наличие связи между формой теста и распределением заданий по уровням DOK; величина эффекта интерпретируется как небольшая/умеренная в зависимости от принятого контекста и размера таблицы сопряженности

Если объединить обе формы, то из 80 заданий 37 (46,25%) классифицированы как DOK-1, 26 (32,50%) - как DOK-2 и 17 (21,25%) - как DOK-3. Все три уровня представлены, однако задания на воспроизведение образуют самую крупную группу.

Наиболее заметное расхождение получено по DOK-1. Во второй форме доля таких заданий выше на 22,5 процентного пункта: 57,5% против 35,0%. Для пилотного сравнения это превышает рабочий ориентир в 10 пунктов. Значение Cramér's $V = 0,23$ указывает, что различие распределений нельзя считать пренебрежимо малым; его необходимо проверить на большем числе форм.

Если объединить DOK-2 и DOK-3, получаем 65,0% в «Химии-1» и 42,5% в «Химии-2». Иными словами, первая форма чаще требует применения знаний или более сложного рассуждения, тогда как во второй заметно больше заданий на воспроизведение.

Важно, что это расхождение возникает при одинаковой тематической структуре. Содержательная матрица и когнитивный профиль рассматриваются и когнитивный профиль как разные стороны сопоставимости. При этом одного DOK также недостаточно, чтобы судить об эквивалентности форм: его нужно рассматривать вместе с содержательными и психометрическими данными.

Содержательный охват мы проверяли отдельно - по числу заданий в каждом основном разделе. Такой контроль показывает, выдержана ли тематическая матрица, но не раскрывает, какие действия выполняет тестируемый внутри этих тем. Такая сопоставимость предполагает согласованность содержательного представления, когнитивных требований, форматов заданий и других характеристик, определенных спецификацией теста.

Таблица 3 – Распределение заданий по содержательным разделам в двух формах теста по химии

Раздел	Номера заданий на форме 1	Число заданий	Номера заданий на форме 2	Число заданий
Частицы вещества	1, 2, 36	3	1, 2, 36	3
Закономерности протекания химических реакций	3, 4, 17	3	3, 4, 17	3
Энергетика химических реакций	5, 37	2	5, 37	2
Химия и жизнь	6-12, 31-33	10	6-12, 31-33	10
Химия вокруг нас	13-16, 34-35, 38-39	8	13-16, 34-35, 38-39	8
Задачи	18-25, 40	9	18-25, 40	9
Контекстные задания	26-30	5	26-30	5

Для параллельных форм ожидается сопоставимое представление содержательных областей, заданных спецификацией. Поэтому первым шагом сравнили фактическое число заданий по разделам.

Установлено, что в обеих формах число заданий совпадает по всем основным разделам (таблица 3). То есть по количественному содержательному охвату «Химия-1» и «Химия-2» различий не показали.

Однако после DOK-кодирования выявлено, что совпадения по темам недостаточно для совпадения когнитивных профилей.

Это сопоставление стало одним из основных результатов: две формы могут содержать одинаковое число заданий по каждой содержательной области и при этом по-разному распределять воспроизведение, применение и стратегическое мышление.

Поэтому при сборке параллельных вариантов целесообразно проверять DOK-профиль отдельно, не подменяя им содержательную экспертизу.

На следующем этапе мы сопоставили категории А, В и С с фактической трудностью и уровнями DOK. В этот анализ вошли все задания обеих форм.

Спецификация задает соотношение 50/30/20 для категорий А, В и С. Для формы из 40 заданий это соответствует 20, 12 и 8 заданиям; именно это соотношение мы использовали как проектную основу.

Чтобы посмотреть, как проектные категории проявляются в данных, использовался показатель p-value. Распределения доли правильных ответов представлены на рисунке 2.

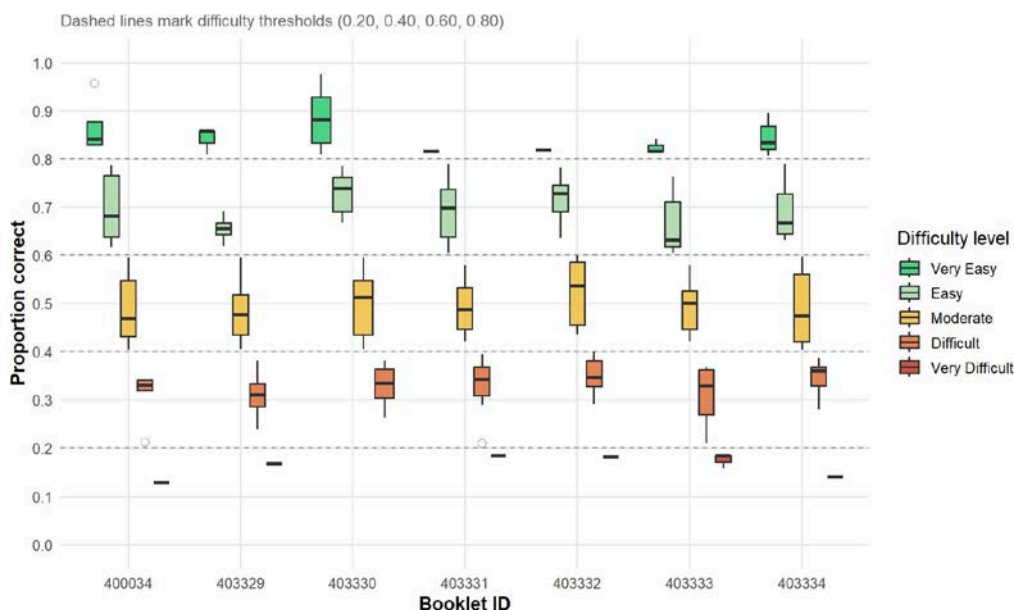


Рисунок 2 – Распределение р-значений заданий (доли правильных ответов) по проектным категориям трудности

Внутри каждой проектной категории мы обнаружили заметный разброс p-value. Распределения соседних категорий перекрываются, особенно А с В и В с С (рисунок 2). Поэтому не представляется возможным связать буквенную категорию с единственным диапазоном наблюдаемой трудности.

Одинаковая пропорция 50/30/20 в двух формах не приводит к одинаковому распределению p value внутри категорий. Следовательно, формальное совпадение проектной структуры само по себе не гарантирует совпадения эмпирической трудности.

При сопоставлении с DOK получена та же картина: проектная категория и когнитивный уровень описывают разные свойства задания. А/В/С фиксирует ожидаемую трудность на этапе разработки, а DOK - глубину действия, необходимого для решения. Поэтому более высокая категория трудности не интерпретируется как автоматический признак более высокого DOK.

При оценке параллельности эти показатели следует рассматривать отдельно: проектное распределение А/В/С, фактическую трудность и DOK-профиль.

Использование стимульного материала. Отдельно рассмотрен тип стимульного материала. В «Химии-1» текстовый или числовой стимул использовался в 60% заданий, в «Химии-2» - в 51%. Без отдельного стимула было 37% и 45% заданий соответственно, графика встречалась редко - в 3% и 4%.

Таблица 4 – Распределение типов стимульного материала в формах теста по химии

Профиль (№ формы)	Есть стимул — графический (% заданий)	Есть стимул — текстовый/числовой (% заданий)	Нет стимула (% заданий)
Химия-1	3	60	37
Химия-2	4	51	45

Если объединить обе формы, наблюдается следующая картина: 55,5% заданий содержат текстовый или числовой стимул, 41,0% не имеют отдельного стимульного материала и 3,5% используют графическое представление.

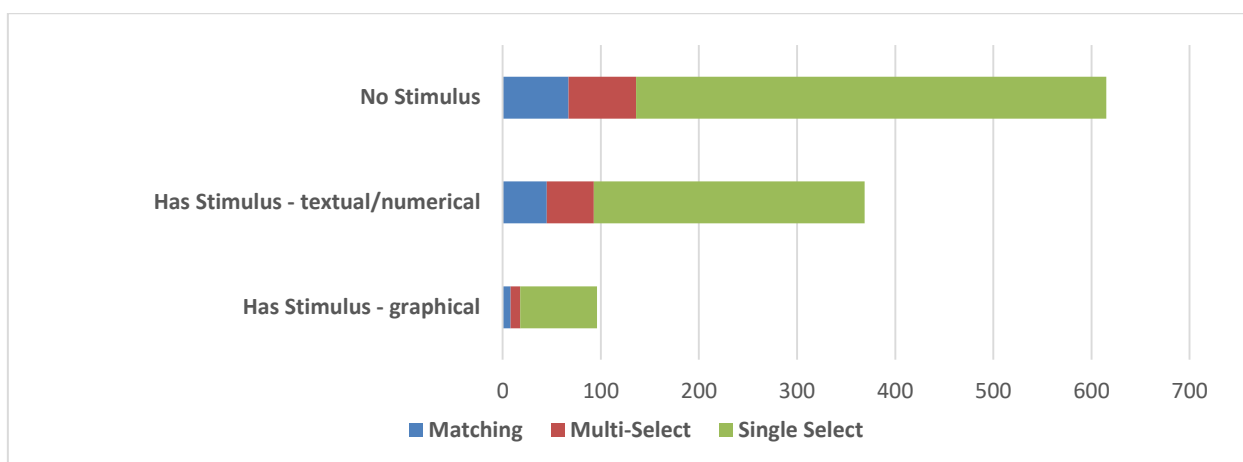


Рисунок 3 – Типы стимульного материала, включенные в задания профильного теста по химии

Мы не связываем преобладание текста и чисел с определенным уровнем ДОК. Один и тот же тип стимула может сопровождать воспроизведение, применение известного способа или стратегическое рассуждение. При кодировании для нас было важно не наличие текста, числа или рисунка как такового, а то, что именно участник должен сделать с этой информацией.

Сопоставление психометрических характеристик двух форм

Психометрический анализ проводился отдельно для каждой формы в Winsteps 5.8.3.0. Для «Химии-1» использовали ответы 121 абитуриента, для «Химии-2» - 161. Эти показатели использовались не для прямого сравнения способностей двух выборок, а для описания поведения заданий и их сопоставления с экспертными ДОК-кодами.

Таблица 5 – Сводные психометрические показатели форм «Химия-1» и «Химия-2»

Показатель	Химия-1 (N=121)	Химия-2 (N=161)
Надежность заданий (Item Reliability)	0.92	0.89
Надежность испытуемых (Person Reliability)	0.85	0.81
Cronbach's α	0.87	0.83
Разделимость заданий (Item Separation)	3.35	2.82
Средняя трудность заданий (logit)	0.00	0.00
Стандартное отклонение трудности заданий	0.71	0.68

Примечание. Средняя трудность заданий установлена равной 0.00 logit в результате центрирования, принятого в модели Рашиа. Стандартное отклонение отражает разброс трудностей относительно этого среднего значения.

По заданиям обе формы показали высокую надежность: item reliability составила 0,92 и 0,89. Person reliability и item separation несколько выше для «Химии-1». Разброс трудностей почти одинаков - 0,71 против 0,68 logit. Поскольку формы выполняли разные группы, эти различия напрямую с ДОК-профилем не связываются. Для более детального сопоставления задания каждой формы сгруппированы по итоговому ДОК и рассчитаны средние значения MEASURE. Результаты приведены в таблице 6.

Таблица 6 – Средняя эмпирическая трудность (logit) заданий по уровням ДОК

Уровень ДОК	Химия-1	Химия-2
ДОК-1	-0.37 (n=14)	-0.15 (n=23)
ДОК-2	0.15 (n=16)	0.08 (n=10)
ДОК-3	0.52 (n=10)	0.35 (n=7)

Примечание. Расчет произведен на основе колонки «MEASURE» из таблиц статистики заданий (TABLE 14.1 из Winsteps 5.8.3.0) для каждой формы. Группировка по уровням ДОК выполнена согласно экспертной классификации.

В обеих формах наблюдается рост средней трудности от DOK-1 к DOK-3: разница между крайними уровнями составляет 0,89 logit для «Химии-1» и 0,50 logit для «Химии-2». Внутри групп сохраняется существенное перекрытие. Например, CONT-26, отнесенное к DOK-3, имеет трудность -0,14 logit, тогда как MC-30 с DOK-1 - 1,55 logit. DOK и эмпирическую трудность нельзя использовать как взаимозаменяемые показатели.

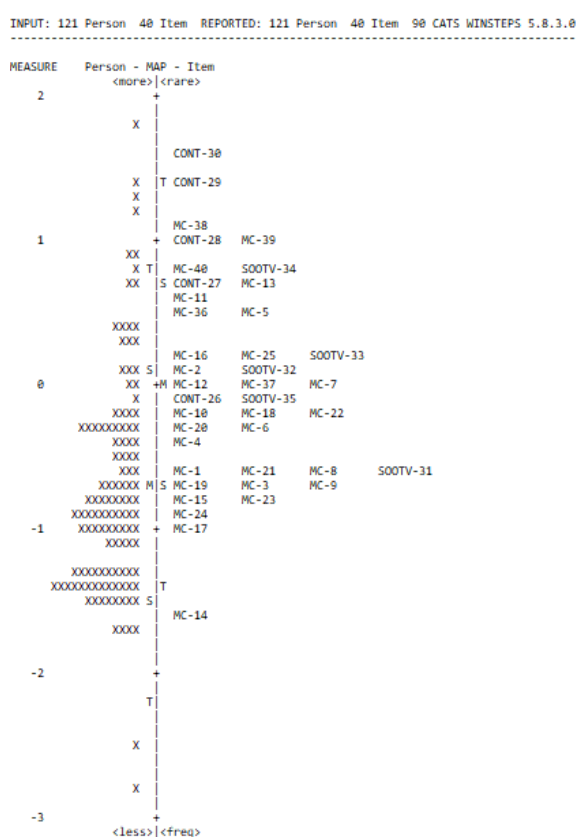


Рисунок 4 – Person - MAP - Item:
Химия-1

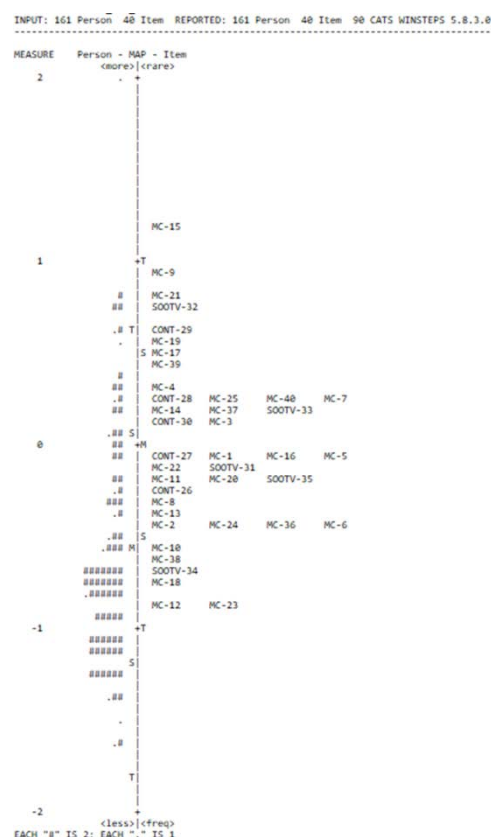


Рисунок 5 – Person - MAP - Item:
Химия-2

Ту же картину мы видим на person-item maps. В «Химии-1» многие задания DOK-3 - CONT-29, CONT-30, MC-38, MC-39, MC-40, SOOTV-34 - располагаются примерно между 0,5 и 1,5 logit, а значительная часть DOK-1 - ниже. Но границы групп пересекаются: MC-30 (DOK-1; 1,55 logit) труднее CONT-26 (DOK-3;-0,14) и SOOTV-35 (DOK-3;-0,15). Во второй форме DOK-3 задания MC-15 (1,19) и MC-9 (0,97 logit) находятся в верхней части шкалы, однако положительные значения есть и у отдельных DOK-1 заданий, например MC-4 (0,31) и CONT-28 (0,28 logit).

По fit-статистикам большинство заданий оказалось в интервале 0,80-1,20 для infit и outfit MNSQ. В «Химии-1» внимание привлекли MC-16 (infit 1,34; ZSTD 3,28) и SOOTV-33 (1,37; 2,75). Во второй форме выделяются SOOTV-31 (1,20; 2,13) и MC-36 (1,18; 1,95). Эти задания

— кандидаты для отдельного содержательного просмотра, прежде всего формулировок и возможных альтернативных стратегий решения. Единичные отклонения не меняют общей картины по формам.

В психометрических данных не обнаружено оснований считать DOK просто отражением трудности. Person reliability составила 0,85 и 0,81, item separation - 3,35 и 2,82, а разброс трудности почти одинаков. Две независимые выборки не позволяют утверждать, будто различия этих показателей вызваны именно составом DOK.

При проектировании параллельных форм целесообразно одновременно контролировать как минимум два аспекта: положение заданий на эмпирической шкале трудности и характер познавательных действий, необходимых для их выполнения.

Обсуждение

Главный результат исследования состоит в том, что две формы, совпадающие по числу заданий в содержательных разделах, не совпали по распределению DOK. Из этого следует ограниченный, но практический вывод: содержательная матрица описывает только одну сторону сопоставимости и не показывает, насколько похожа глубина работы с предметным материалом.

Наличие в обеих формах DOK-2 и DOK-3 рассматривается в контексте международного развития естественно-научного оценивания, где наряду с воспроизведением проверяются применение знаний, интерпретация данных, объяснение явлений и работа с моделями. Такие требования представлены в PISA, AP Chemistry и IB Chemistry [16, 17, 18]. В казахстанских исследованиях химического образования также обсуждается акцент на предметных компетенциях [19].

Второй результат - отсутствие прямого совпадения DOK с проектной категорией A/B/C и фактической трудностью. Перекрытие p-value между категориями и случаи, когда отдельное задание DOK-1 оказывается труднее DOK-3, показывают, что успешность решения зависит не только от глубины рассуждения. На результат могут влиять контекст, формулировка, объем информации и способ предъявления. Поэтому проектная трудность не является заменой когнитивному кодированию.

Анализ стимульного материала приводит к сходному выводу. Текст, числовые данные или графика сами по себе не повышают DOK; важен способ работы с ними. Для химии это особенно существенно, поскольку решение может требовать перехода между наблюдаемыми явлениями, субмикроскопическими представлениями и символическим языком [20, 21].

С практической точки зрения не предлагается просто увеличивать долю DOK-3. Более продуктивным является заранее задавать требуемое сочетание воспроизведения, применения и стратегического рассуждения и контролировать его вместе с тематической матрицей и трудностью. В таком случае DOK может стать дополнительным параметром спецификации, а не конкурирующей системой оценки качества.

Заключение

В нашем анализе две формы совпадали по количеству заданий в содержательных разделах, но различались по когнитивному профилю. Cramér's V для связи формы с уровнем DOK составил 0,23. Наиболее крупное различие мы получили по DOK-1: 57,5% во второй форме и 35,0% в первой, то есть 22,5 процентного пункта. Для пилотного сравнения это выше нашего рабочего ориентира в 10 пунктов. При этом мы не используем ни этот ориентир, ни Cramér's V как самостоятельное основание объявлять формы непараллельными, поскольку исследование охватывает только два варианта.

Полученные данные показывают, что показывают ограниченность оценки формы только по содержательной матрице и проектному соотношению категорий трудности.

Дополнительный DOK профиль помогает увидеть, какие типы познавательных действий представлены в форме и в какой пропорции.

Следующим этапом планируется включить в анализ большее число форм ЕНТ по химии. Это позволит проверить устойчивость найденных различий и перейти от пилотного описания к обоснованию целевого когнитивного профиля. Отдельно мы рассматриваем возможность проведения когнитивных интервью с учащимися, поскольку именно такой метод позволит сопоставить экспертно присвоенный DOK с фактическими стратегиями решения.

Список литературы

1. Messick S. Validity // Educational measurement / Ed. by R.L. Linn. - 3rd ed. - New York: American Council on Education / Macmillan, 1989. - P. 13-103.
2. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing. - Washington, DC: AERA, 2014. - 230 p.
3. Gorin J.S., Embretson S.E. Item difficulty modeling of paragraph comprehension items // Applied Psychological Measurement. - 2006. - Vol. 30, No. 5. - P. 394-411. <https://doi.org/10.1177/0146621606288554>
4. Shavelson R.J. On the measurement of competency // Empirical Research in Vocational Education and Training. - 2010. - Vol. 2, No. 1. - P. 41-63. <https://doi.org/10.1007/BF03546488>
5. Ravand H., Shahi R., Effatpanah F., Moghadamzadeh A. Measuring cognitive levels in high-stakes testing: A CDM analysis of a university entrance examination using Bloom's taxonomy // Frontiers in Education. - 2025. - Vol. 10. - Art. 1498765. <https://doi.org/10.3389/feduc.2025.1644811>
6. Webb N.L. Criteria for alignment of expectations and assessments in mathematics and science education (Research Monograph No. 6). - Madison, WI: Wisconsin Center for Education Research, 1997. - 45 p.
7. Webb N.L. Alignment, depth of knowledge, and change // Presentation at the 50th Annual Meeting of the Florida Educational Research Association. - Miami, FL, 2005. - 12 p.
8. Holme T., Knaus K., Murphy K., Blecking A. A valid and reliable instrument for cognitive complexity rating assignment of chemistry exam items // Journal of Chemical Education. - 2011. - Vol. 88, No. 5. - P. 554-560. <https://doi.org/10.1021/ed900070y>
9. Raker J.R., Trate J.M., Holme T.A., Murphy K. Adaptation of an instrument for measuring the cognitive complexity of organic chemistry exam items // Journal of Chemical Education. - 2013. - Vol. 90, No. 10. - P. 1290-1295. <https://doi.org/10.1021/ed400373c>
10. Национальный центр тестирования. Спецификация теста по химии для Единого национального тестирования. - Астана: НЦТ МНВО РК, 2025. - 12 с.
11. Садыков Т.М., Кокибасова Г.Т., Нурмыш С.Д. Development of multi-level tasks for the development of natural science literacy of students in chemistry // Gumilyov Journal of Pedagogy. - 2025. - Vol. 150, No. 1. - P. 42-57. <https://doi.org/10.32523/3080-1710-2025-150-1-42-57>
12. Шаңытбаева Ж.Н. Формирование химических знаний у обучающихся на основе познавательных заданий по химии // Yessenov Science Journal. - 2026. - № 1 (54). - С. 149-156. - <https://doi.org/10.56525/km8rza83>
13. Okanlawon A.E., Adeoti Y.F. Content analysis of West African senior school certificate chemistry examination questions according to cognitive complexity // IFE PsychologIA. - 2014. - Vol. 22, No. 2. - P. 1-12.
14. Zorluoğlu S.L., Kızılaslan A., Dönmez Yapıcıoğlu M. The analysis of 9th grade chemistry curriculum and textbook according to revised Bloom's taxonomy // Cypriot Journal of Educational Sciences. - 2020. - Vol. 15, No. 1. - P. 9-20. <https://doi.org/10.18844/cjes.v15i1.3516>
15. Ren D., Hu T., Dong X., Yao H. Assessing chemistry disciplinary competence in high-stakes examination forms: Evidence from the 2023 Chinese Gaokao // Frontiers in Education. - 2026. - Vol. 11. - Art. 1884670. <https://doi.org/10.3389/feduc.2026.1884670>
16. OECD. PISA 2025 science framework. - Paris: OECD Publishing, 2025. - 180 p.
17. College Board. AP Chemistry course and exam description (Effective Fall 2024). - New York: College Board, 2024. - 240 p.

18. International Baccalaureate Organization. Diploma Programme chemistry guide (First assessment 2025). - Cardiff: IBO, 2023. - 112 p.
19. Alzhanova G., Altynbayeva L., Zhanbulatova G., Tleulessova A., Toltebayeva F. Exploring competency-based chemistry teaching: Trends, challenges and perspectives // Chemistry-Didactics-Ecology-Metrology. - 2025. - Vol. 30, No. 1-2. - P. 87-97. <https://doi.org/10.2478/cdem-2025-0008>
20. Johnstone A.H. Why is science difficult to learn? Things are seldom what they seem // Journal of Computer Assisted Learning. - 1991. - Vol. 7, No. 2. - P. 75-83. <https://doi.org/10.1111/j.1365-2729.1991.tb00230.x>
21. Multiple representations in chemical education / Ed. by J.K. Gilbert, D. Treagust. - Dordrecht: Springer, 2009. - 439 p.
22. Wright B.D., Stone M.H. Best test design. - Chicago: MESA Press, 1979. - 238 p.
23. Linacre J.M. Sample size and item calibration stability // Rasch Measurement Transactions. - 1994. - Vol. 7, No. 4. - P. 328-332.
24. Smith A.B., Rush R., Fallowfield L.J., Velikova G., Sharpe M. Rasch fit statistics and sample size considerations for polytomous data // BMC Medical Research Methodology. - 2008. - Vol. 8, No. 1. - Art. 33. <https://doi.org/10.1186/1471-2288-8-33>

Ш.Б. Алтыбаева, А.Н. Қуанышбекова, Ж. Абдулла, А.Н. Есимбекова

ҰЛТТЫҚ БІРЫҢҒАЙ ТЕСТІЛЕУДЕГІ ХИМИЯ ТАПСЫРМАЛАРЫНЫҢ КОГНИТИВТІК КҮРДЕЛІЛІГІН БАҒАЛАУ: ҚАЗАҚСТАНДАҒЫ ҰБТ ДЕРЕКТЕРІ

Ірі ауқымды қабылдау емтихандарында екі нұсқаның бірдей тақырыптарды қамтуы олардың танымдық талаптары да бірдей дегенді білдірмейді. Осы пилоттық зерттеуде біз Қазақстанның ҰБТ химия пәні бойынша екі нұсқасындағы 80 тапсырманы талдадық. Тапсырмалардың когнитивтік күрделілігін Уэббтің Depth of Knowledge (DOK) моделінің үш деңгейлі бейімделуі арқылы бағаладық. Алты сарапшы барлық тапсырманы жеке кодтады; бастапқы абсолюттік келісім 87% болды, ал келіспеген жағдайларды кейін ортақ талқылауда нақтыладық. «Химия-1» нұсқасында DOK-1, DOK-2 және DOK-3 үлестері 35,0%, 40,0% және 25,0% болса, «Химия-2» нұсқасында 57,5%, 25,0% және 17,5% болды. Ең үлкен айырманы DOK-1 бойынша байқадық - 22,5 пайыздық тармақ. Раш моделі бойынша орташа эмпирикалық қиындық DOK-1-ден DOK-3-ке қарай өсті, бірақ деңгейлер арасындағы қиындық диапазоңдары толық бөлінген жоқ. Осы нәтижелерге сүйене отырып, біз мазмұндық бөлімдер бойынша тапсырмалар санының сәйкес келуі когнитивтік талаптардың да сәйкес болуына кепілдік бермейді деп есептейміз. DOK-профильді параллель нұсқаларды құрастыру мен сараптауда қосымша сипаттама ретінде пайдалануға болады. Зерттеу екі нұсқамен шектелгендіктен, бұл қорытындыларды кеңірек материалда тексеру қажет.

Түйін сөздер: когнитивтік күрделілік, DOK, ҰБТ, химия, жоғары ставкалы тестілеу, валидтілік, Раш моделі, нысандардың параллельдігі.

S.B. Altybayeva, A.N. Kuanyshbekova, Z. Abdulla, A.N. Yessimbekova

ASSESSMENT OF THE COGNITIVE COMPLEXITY OF CHEMISTRY TASKS IN THE NATIONAL ENTRANCE EXAMINATION: EVIDENCE FROM KAZAKHSTAN'S UNT

Two entrance-test forms may cover the same content while placing different cognitive demands on examinees. In this pilot study, we examined 80 chemistry items from two forms of Kazakhstan's Unified National Testing (UNT). We classified item demands using a three-level adaptation of Webb's Depth of Knowledge (DOK) framework. Six experts coded all items independently. Initial absolute agreement was 87%, and we resolved disagreements through subsequent discussion. In Chemistry-1, we classified 35.0% of the items as DOK-1, 40.0% as DOK-2, and 25.0% as DOK-3; in Chemistry-2, the corresponding proportions were 57.5%, 25.0%, and 17.5%. The largest difference was observed for DOK-1 and reached 22.5 percentage points. Our Rasch analysis showed an increase in mean empirical difficulty from DOK-1 to DOK-3, although the difficulty ranges overlapped. We therefore conclude that matching the number of items

across content areas does not necessarily produce matching cognitive profiles. We view DOK as an additional descriptor of parallel forms alongside content and psychometric information. Because our analysis covered only two forms, these findings should be treated as preliminary.

Keywords: cognitive complexity, DOK, UNT, chemistry, high-stakes testing, validity, Rasch model, test form parallelism.

References

1. Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education / Macmillan.
2. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (p. 230). AERA.
3. Gorin, J. S., & Embretson, S. E. (2006). Item difficulty modeling of paragraph comprehension items. *Applied Psychological Measurement*, 30(5), 394–411. <https://doi.org/10.1177/0146621606288554>
4. Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, 2(1), 41–63. <https://doi.org/10.1007/BF03546488>
5. Ravand, H., Shahi, R., Effatpanah, F., & Moghadamzadeh, A. (2025). Measuring cognitive levels in high-stakes testing: A CDM analysis of a university entrance examination using Bloom's taxonomy. *Frontiers in Education*, 10, Article 1498765. <https://doi.org/10.3389/feduc.2025.1644811>
6. Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6, p. 45). Wisconsin Center for Education Research.
7. Webb, N. L. (2005, November). *Alignment, depth of knowledge, and change* (p. 12) [Paper presentation]. 50th Annual Meeting of the Florida Educational Research Association, Miami, FL, United States.
8. Holme, T., Knaus, K., Murphy, K., & Blecking, A. (2011). A valid and reliable instrument for cognitive complexity rating assignment of chemistry exam items. *Journal of Chemical Education*, 88(5), 554–560. <https://doi.org/10.1021/ed900070y>
9. Raker, J. R., Trate, J. M., Holme, T. A., & Murphy, K. (2013). Adaptation of an instrument for measuring the cognitive complexity of organic chemistry exam items. *Journal of Chemical Education*, 90(10), 1290–1295. <https://doi.org/10.1021/ed400373c>
10. Natsional'nyy tsentr testirovaniya. (2025). *Spetsifikatsiya testa po khimii dlya Yedinogo natsional'nogo testirovaniya* [Chemistry test specification for the Unified National Testing] (p. 12). NTsT MNVO RK.
11. Sadykov, T. M., Kokibasova, G. T., & Nurmysh, S. D. (2025). Development of multi-level tasks for the development of natural science literacy of students in chemistry. *Gumilyov Journal of Pedagogy*, 150(1), 42–57. <https://doi.org/10.32523/3080-1710-2025-150-1-42-57>
12. Shañytbaeva, Zh. N. (2026). Formirovanie khimicheskikh znaniy u obuchayushchikhsya na osnove poznavatel'nykh zadaniy po khimii [Formation of chemical knowledge in students based on cognitive chemistry tasks]. *Yessenov Science Journal*, 1(54), 149–156. <https://doi.org/10.56525/km8rza83>
13. Okanlawon, A. E., & Adeoti, Y. F. (2014). Content analysis of West African senior school certificate chemistry examination questions according to cognitive complexity. *IFE Psychologia*, 22(2), 1–12.
14. Zorluoğlu, S. L., Kızılaslan, A., & Dönmez Yapıcıoğlu, M. (2020). The analysis of 9th grade chemistry curriculum and textbook according to revised Bloom's taxonomy. *Cypriot Journal of Educational Sciences*, 15(1), 9–20. <https://doi.org/10.18844/cjes.v15i1.3516>
15. Ren, D., Hu, T., Dong, X., & Yao, H. (2026). Assessing chemistry disciplinary competence in high-stakes examination forms: Evidence from the 2023 Chinese Gaokao. *Frontiers in Education*, 11, Article 1884670. <https://doi.org/10.3389/feduc.2026.1884670>
16. OECD. (2025). *PISA 2025 science framework* (p. 180). OECD Publishing.
17. College Board. (2024). *AP Chemistry course and exam description (Effective Fall 2024)* (p. 240). College Board.
18. International Baccalaureate Organization. (2023). *Diploma Programme chemistry guide (First assessment 2025)* (p. 112). IBO.

19. Alzhanova, G., Altynbayeva, L., Zhanbulatova, G., Tleulessova, A., & Toltebayeva, F. (2025). Exploring competency-based chemistry teaching: Trends, challenges and perspectives. *Chemistry-Didactics-Ecology-Metrology*, 30(1–2), 87–97. <https://doi.org/10.2478/cdem-2025-0008>
20. Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, 7(2), 75–83. <https://doi.org/10.1111/j.1365-2729.1991.tb00230.x>
21. Gilbert, J. K., & Treagust, D. (Eds.). (2009). *Multiple representations in chemical education* (p. 439). Springer.
22. Wright, B. D., & Stone, M. H. (1979). *Best test design* (p. 238). MESA Press.
23. Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4), 328–332.
24. Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8(1), Article 33. <https://doi.org/10.1186/1471-2288-8-33>

Сведения об авторах:

Алтыбаева Шугыла Болатовна (автор-корреспондент) - РГП на ПХВ «Национальный центр тестирования» МНВО РК, руководитель управления, Астана, Казахстан, e-mail: shygyla.altybayeva@gmail.com

Куанышбекова Арайлым Ниязбекқызы - РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий эксперт, Астана, Казахстан, e-mail: quanyshbekova98@mail.ru

Абдулла Жанар - КГУ «Специализированный лицей № 82 «Дарын», учитель химии, Астана, Казахстан, e-mail: jeneiro@mail.ru

Есимбекова Аруна Нурлыбаевна - РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий эксперт, Астана, Казахстан, e-mail: aruna.iesimbekova@mail.ru

Авторлар туралы мәлімет:

Алтыбаева Шугыла Болатовна (автор-корреспондент) - ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, бөлім басшысы, Астана, Қазақстан, e-mail: shygyla.altybayeva@gmail.com

Куанышбекова Арайлым Ниязбекқызы - ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, жетекші сарапшы, Астана, Қазақстан, e-mail: quanyshbekova98@mail.ru

Абдулла Жанар - «№ 82 «Дарын» мамандандырылған лицейі» КММ, Химия пәні мұғалімі, Астана, Қазақстан, e-mail: jeneiro@mail.ru

Есимбекова Аруна Нурлыбаевна - ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, жетекші сарапшы, Астана, Қазақстан, e-mail: aruna.iesimbekova@mail.ru

Information about the Authors:

Altybayeva Shugyla (corresponding author) - Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Head of Department, Astana, Kazakhstan, e-mail: shygyla.altybayeva@gmail.com

Kuanyshbekova Arailym - Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Leading Expert, Astana, Kazakhstan, e-mail: quanyshbekova98@mail.ru

Abdulla Zhanar - KSU «Specialized Lyceum No. 82 «Daryn», Chemistry Teacher, Astana, Kazakhstan, e-mail: jeneiro@mail.ru

Yessimbekova Aruna - Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Leading Expert, Astana, Kazakhstan, e-mail: aruna.iesimbekova@mail.ru

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖӘНЕ БІЛІМ БЕРУДІ БАҒАЛАУДАҒЫ ИННОВАЦИЯ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ИННОВАЦИИ В ОЦЕНИВАНИИ ОБРАЗОВАНИЯ

ARTIFICIAL INTELLIGENCE AND INNOVATIONS IN EDUCATIONAL ASSESSMENT

А.Кушубек^{1*}, Б.Шарипбеков², Б.Рахимов³, С.Сантеева⁴, А.Шайханова⁵
^{1,2,3,4,5}ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана қ, Қазақстан Республикасы
*e-mail: b.sharipbekov@testcenter.kz

¹ORCID 0009-0006-2436-314X, ²ORCID 0009-0006-4551-1957, ³ORCID 0009-0001-3609-960X

⁴ORCID 0000-0001-9426-6704, ⁵ORCID 0000-0001-6006-4813

ІТ-ИНФРАҚҰРЫЛЫМДАРДЫҢ АҚПАРАТТЫҚ ҚАУІПСІЗДІГІН БАСҚАРУДА ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕГІ ЛОГ ТАЛДАУ ЖӘНЕ АВТОМАТТАНДЫРЫЛҒАН ӘРЕКЕТ ЕТУ АРХИТЕКТУРАЛАРЫНА ШОЛУ

Соңғы он жылда киберқауіпсіздік саласы елеулі өзгерістерге ұшырады: технологиялардың қарқынды дамуымен қатар жаңа қатерлердің саны да артып келеді. Ұйымдар инсайдерлік қауіптерге, онлайн тыңшылыққа, зиянды бағдарламаларға және деректердің сыртқа ағуына бір мезгілде тап болуда. Кәсіпорын қызметі өзара байланысқан ақпараттық жүйелерге сүйенетіндіктен, кибершабуыл бүкіл жұмыс процесін бұзып, айтарлықтай қаржылық шығынға әкелуі мүмкін. Осыған байланысты компаниялар құпия деректері мен сандық меншігін қорғауда алдын алу шараларына көбірек көңіл бөлуде. Мақалада жасанды интеллект технологияларын пайдалана отырып, ІТ инфрақұрылымның ақпараттық қауіпсіздігін орталықтандырып басқару әдістері мен архитектурасы қарастырылады. Қауіпсіздік оқиғаларын жинап, өзара байланыстыру үшін SIEM жүйелерін қолдануға, сондай-ақ аномалияларды анықтап, оқиғаларды талдайтын интеллектуалды алгоритмдерді біріктіруге басты назар аударылады. Оқиғаларды басқару жүйесін, жасанды интеллект модулін және автоматтандырылған жауап беруге арналған SOAR платформасын біріктіретін архитектура жобасы ұсынылады. Журнал талдауына негізделген қатерлерді анықтау мен азайтудың практикалық сценарийлері, қорғаныс шараларын автоматтандыру мүмкіндіктері талданады. Нәтижелер интеллектуалды талдау мен процестерді автоматтандыру оқиғаларды анықтау тиімділігін арттырып, ақпараттық қауіпсіздік мамандарының жұмыс жүктемесін азайтып, жауап беру уақытын қысқартатынын көрсетеді.

Түйін сөздер: киберқауіп шабуылдар, ақпараттық қауіпсіздік, SIEM, жасанды интеллект, SOAR, автоматтандыру, қауіпсіздік оқиғалары.

Кіріспе

Технологиялардың үздіксіз дамуы жаңа киберқауіптердің пайда болуын күн санап арттырып келеді. Осының салдарынан киберқауіпсіздік саласы соңғы онжылдықта түбегейлі өзгерді. Бүгінде әрбір ұйым бірнеше тәуекелге бір мезгілде тап болады: инсайдерлік қауіптер, онлайн тыңшылық, компьютерлік вирустар және деректердің рұқсатсыз сыртқа ағуы. Кәсіпорындардың қызметі технологиялар мен өзара байланысқан ақпараттық жүйелерге тәуелді болғандықтан, кибершабуыл ықтималдығы да, оның операциялық қызмет пен қаржылық нәтижелерге тигізетін әсері де арта түсуде. Мұндай жағдай тиімді қорғаныс шараларын талап етеді - сондықтан кәсіпорындар құпия деректері мен сандық меншігін қорғауға белсенді кірісуде [1].

Бұл зерттеу SIEM, SOAR және жасанды интеллект технологияларын қамтитын заманауи ақпараттық қауіпсіздік жүйелерін кешенді талдауды мақсат етеді: олардың негізгі сипаттамалары, артықшылықтары, енгізудің озық тәжірибелері мен қолданыстағы шектеулері қарастырылады. Мақалада ұйымдарды заманауи киберқауіптерден қорғауды күшейту үшін SIEM және SOAR шешімдерін жасанды интеллектпен біріктірудің маңызы зерделенеді. Осы жүйелердің жалпы киберқауіпсіздік архитектурасындағы рөлі, сондай-ақ мониторинг, оқиғаларды корреляциялау және оларға жауап беру процестерін автоматтандыру мүмкіндіктері талданады.

Ұйым инфрақұрылымын кибершабуылдардан қорғау бүгінде әлемдік ауқымдағы басым міндетке айналды. Қатерлер күрделене түскен сайын, дәстүрлі қауіпсіздік шараларының жеткіліксіз болатыны айқындала бастады.

Зерттеу материалдары мен әдістері

Жасанды интеллект (AI) пен машиналық оқытуды (ML) қолдану қауіптерді ерте кезеңде анықтауға мүмкіндік береді әрі анықтау мен жауап беру жүйелерінің мүмкіндіктерін едәуір кеңейтеді. Жүйе архитектурасы бірнеше ақпараттық технологияны өзара біріктіруге негізделген. SIEM (Security Information and Event Management) жүйелері желіден, серверлерден және цифрлық құрылғылардан (endpoint) журналдарды орталықтандырып жинап, талдау үшін қолданылады. SOAR (Security Orchestration, Automation and Response) инциденттерге автоматтандырылған жауап беру әрекеттерін орындайды. Жасанды интеллект алгоритмдері SIEM арқылы анықталған инциденттерге сұрау салып талдау жасайды және қолданушыға түсінікті жауап қалыптастырады [2]. Талдау мен жауап беру процесі бірнеше кезеңнен тұрады.

Нәтижелер және талқылау

Деректерді жинау. SIEM жүйелері цифрлық инфрақұрылымның барлық деңгейінен оқиғаларды жинайды - желілік белсенділіктен бастап инфрақұрылым жүйелерінің қауіпсіздік журналдарына дейін.

Нақты уақыттағы талдау. Жасанды интеллект компоненттері кіріс деректерін өңдеп, қауіп-қатер мен ауытқуларды алдын ала болжайтын тенденцияларды іздейді.

Автоматтандырылған жауап беру. SOAR платформалары алдын ала жазылған сценарийлерге (playbooks) сәйкес әрекет етеді - мысалы, бұзылған жүйелерді оқшаулайды немесе зиянды IP мекенжайларын блоктайды. Сонымен қатар олар жүйе компоненттері арасындағы байланысты қамтамасыз етеді.

Жасанды интеллект қауіпсіздік сценарийлерін - фишинг, зиянды бағдарлама, инсайдерлік қауіп және DDoS шабуылдарын - модельдеу тиімділігін айтарлықтай арттырды. Анықтау дәлдігі белгілі қауіптер үшін 95%-ға, ал белгісіз (нөлдік күн) шабуылдар үшін 87%-ға жетеді. Жауап беру жылдамдығы да өзгерді: SOAR-да AI қолдану оқиғаға жауап беру уақытын орта есеппен 60%-ға қысқартып, әрекетті бірнеше секунд ішінде орындауға мүмкіндік береді. Бұған қоса, модельдердің дәлдігі жалған дабылдарды шамамен 40%-ға азайтады — бұл аналитиктерді біртекті тексерістен босатып, күрделірек міндеттерге көңіл бөлуге мүрсат береді [2].

Осы артықшылықтарға қарамастан, мұндай жүйелерді енгізу бірқатар қиындыққа тап болады. Біріншіден, деректер сапасы: жүйенің тиімділігі векторлық базаға енгізілген деректердің сапасына тікелей тәуелді, ал дұрыс форматталмаған деректер қауіпке әрекет ету ықтималдығын төмендетуі мүмкін. Екіншіден, жасанды интеллектке бағытталған қарсыласу шабуылдары (adversarial attacks): шабуылдаушылар арнайы дайындалған деректерді енгізіп, AI алгоритмдерін алдауға тырысуы ықтимал. Үшіншіден, ресурс сыйымдылығы: күрделі модельдерді оқыту мен орналастыру елеулі есептеу қуатын талап етеді.

Осы мәселелерді шешу үшін жобада жүктемелерді бөліп өңдеу, деректерді кеңейту әдістерін (data augmentation) енгізу және манипуляцияға төзімді «тұрақты» модельдер жобалау ұсынылады. Мұндай тәсіл ақпараттық инфрақұрылымның шабуылдарға жалпы төзімділігін арттырып қана қоймайды, ұйымдарға үздіксіз өзгеріп отыратын киберқауіпсіздік ландшафтына бейімделуге де жол ашады [2].

Аталған жұмыстарды жүргізудегі өзекті мәселелердің бірі — журнал деректерін талдау. Журнал көздері әртүрлі болуы мүмкін: жұмыс станциялары, серверлер, firewall-дар, желілік құрылғылар, түрлі ақпараттық жүйелер және басқалары. Әр жүйенің журналы өзіндік

форматта жазылатындықтан, талдау алдында оны алдымен өңдеп, содан кейін деректерді векторлық базаға енгізу қажет.

Дәстүрлі талдаудың негізгі мәселелері:

1. Алдын ала өңдеудің күрделілігі. Көптеген жүйе көп сатылы өңдеуді қажет етеді. Мұнда талдау кезеңін - өңделмеген журналдарды құрылымдық деректерге түрлендіруді - дұрыс жасамау қателіктер мен ақпарат жоғалуына әкеледі [3].

2. Шабуыл деректерінің тапшылығы. Нақты жағдайда жүйенің қалыпты жұмысы туралы дерек көп, ал аномалиялар мен шабуыл мысалдары сирек кездеседі. Осы теңгерімсіздік инцидентті болған сәтте талдауды қиындатады [3].

3. Үздіксіз оқытудың болмауы. Қолданыстағы алгоритмдер сарапшылар жаңа қауіп түрлерін анықтаған кезде жеткілікті жылдам жаңартылмайды - бұл олқылық жаңа қатерлерге осал тұс қалдырады [3].

Жоғарыда аталған мәселелерді шешу үшін ұсынылып жүрген тәсілдермен таныстым. Ұсынылатын модель - журнал талдауының барлық кезеңін бір жерде орындайтын біртұтас жүйе. Ол бірнеше маңызды бөліктен құралады.

Күрделі журнал талдауының орнына модель «шикі» деректермен тікелей жұмыс істейді: таңбаларды тікелей сандық форматқа түрлендіреді. Бұл жүйе жұмысын жеңілдетіп қана қоймай, талдау барысында жиі кездесетін қателерді де болдырмайды. Модельдің өзегі - трансформатор. Ол журнал жолдарын талдап, олардың құрылымын түсінеді әрі мәтін бөліктері арасындағы тәуелділіктерді табады. Зейін механизмі оған журналдың қай бөлігі маңызды екенін ажыратуға көмектеседі. Ең соңында қарапайым жіктеуіш түр - ол логтың қалыпты ма, әлде аномалия ма екенін шешеді.

Модель екі кезеңде оқытылады. Біріншісі - қалыпты журналдарда өзін-өзі оқыту. Мұнда модель тек әдеттегі журналдармен жұмыс істейді, шабуыл имитацияланбайды. Ауытқуларды тани білуі үшін қалыпты журналдарға арнайы өзгеріс енгізіледі: таңбалар қосылады, ауыстырылады немесе жойылады. Осылайша модель нендей журналдардың ерекше саналатынын үйренеді. Екіншісі - кері байланыс арқылы қосымша оқыту. Мамандар нақты шабуылдарды тауып, оларды белгілегенде модель толық оқытылады. Нәтижесінде инциденттерді анықтау дәлдігі артады, ал модель өз қателерінен сабақ алып, уақыт өте жетіле түседі [3].

Мұндай тәсіл әртүрлі журнал көздері бар ортада талдау жүйесін құрудың басты кедергісін шешеді — күрделі алдын ала өңдеуді де, шабуыл туралы белгіленген деректердің үлкен көлемін де қажет етпейді. Сонымен қатар ол қауіпсіздік жүйесіне өзгермелі қауіп ландшафтымен бірге дамуға мүмкіндік беріп, икемділікті қамтамасыз етеді.

Қауіпсіздікті басқарудың тиімді архитектурасы бүгінде үш компонентке негізделеді [4] (1-сурет):

1. SIEM (Security Information and Event Management) - бүкіл инфрақұрылымнан оқиғаларды нақты уақытта орталықтандырып жинап, корреляциялайтын жүйе.

2. Жасанды интеллект (AI) - классикалық әдістер таба алмайтын жасырын үлгілер мен ауытқуларды анықтау үшін терең оқыту алгоритмдерін қолданатын компонент.

3. SOAR (Security Orchestration, Automation and Response) - жауап беру сценарийлерін (playbooks) автоматты орындап, қауіпсіздік маманының күнделікті операцияларға араласуын азайтады.

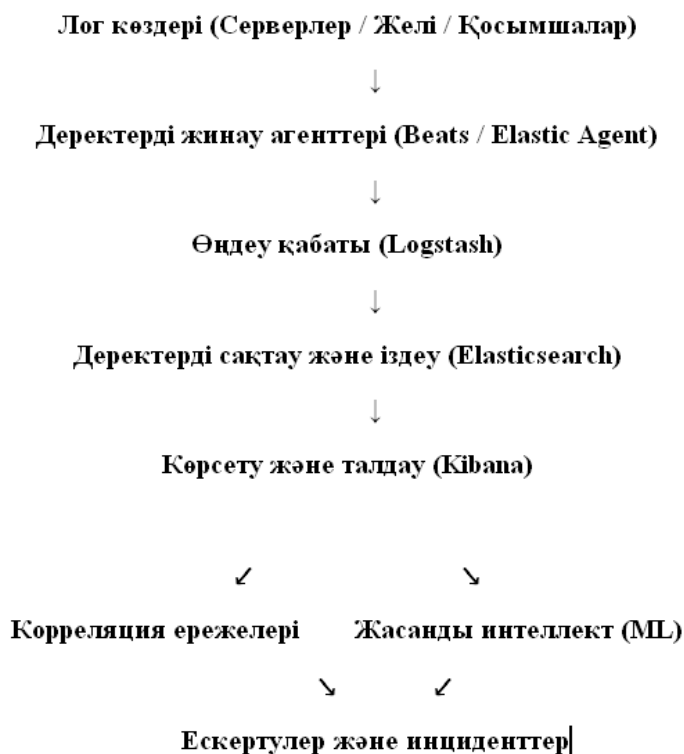
Кең тараған шешімдердің бірі - ELK-стекке негізделген Elastic Security платформасы (1-сурет). Ол журналдарды орталықтандырып жинауды, сақтауды және талдауды қамтамасыз етеді, әрі аномалияларды анықтау үшін кіріктірілген AI механизмдерін пайдаланады [5]. Бұл тәсіл жобада ұсынылған архитектураға жақын: онда да орталықтандырылған журнал талдауы мен AI компоненті қатар қолданылады.

Оның жұмысы келесідей:

1. Логтарды орталықтандырылған жинау (Beats / Agent)
2. Elasticsearch-та сақтау және өңдеу
3. Kibana арқылы визуализация

4. Ережелер мен кірістірілген машиналық оқыту негізінде талдау

Ерекшелігі – кірістірілген машиналық оқыту механизмдері қолданылса, шектеуі – AI алдын ала анықталған алгоритмдермен шектелген.

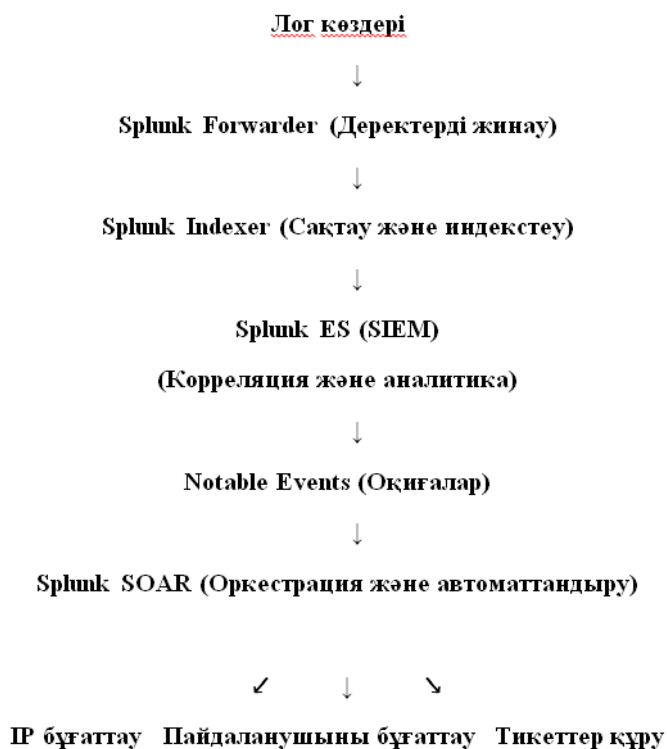


1-сурет – ELK-стек архитектурасы

Кең қолданыстағы тағы бір шешім - SIEM мен SOAR интеграциясын жүзеге асыратын Splunk Enterprise Security және Splunk SOAR біріктірілген жүйесі (2-сурет). Ол қауіпсіздік оқиғаларын талдаумен шектелмейді: playbooks арқылы оқиғаларға автоматты түрде жауап та береді [6]. Дәл осы мүмкіндік жобада ұсынылған жүйенің автоматтандырылған жауап беру тұжырымдамасына толық сәйкес келеді. Жұмыс принципі мынадай:

1. Логтар Splunk Universal Forwarder арқылы жиналып, Indexer-ге жіберіледі.
2. Деректер Parsing және Normalization (CIM – Common Information Model) арқылы бір форматқа келтіріледі.
3. Кореляция механизмдері күдікті әрекеттерді анықтайды. Нәтижесінде инциденттер қалыптасады. Инциденттер Splunk SOAR-ға беріледі. SOAR ішінде playbooks (алдын ала жазылған конфигурациялар) арқылы автоматты әрекеттер орындалады.

Ерекшелігі – икемді корреляция және қуатты автоматтандыру жүйесін қолдануда болса, кемшілігі – AI негізінен статистика мен алдын ала модельдерге тәуелді болады.



2-сурет – Splunk Enterprise Security және Splunk SOAR

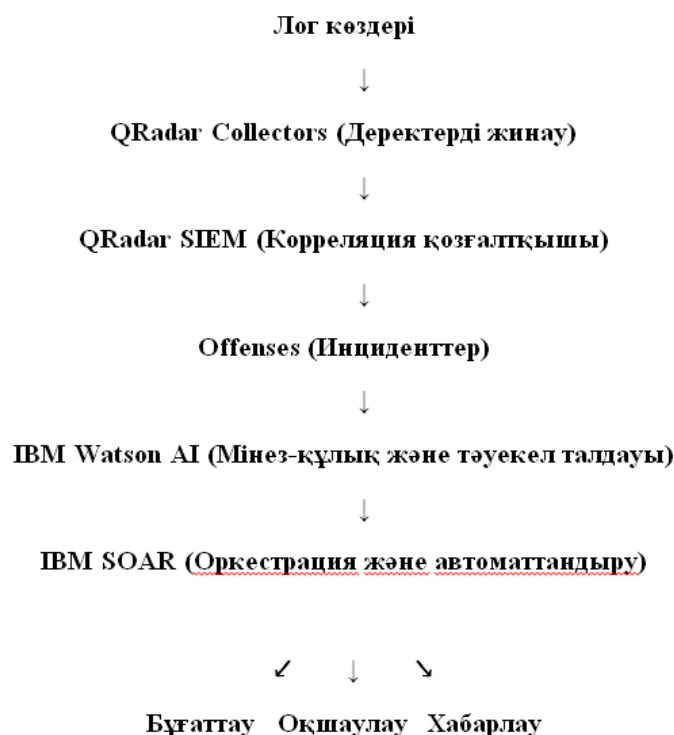
IBM QRadar жүйесі IBM SOAR-мен біріктіріліп, SOC құруға арналған корпоративтік шешім ретінде ұсынылады (3-сурет). Ол қауіпті анықтау дәлдігін арттыру үшін оқиғалар корреляциясын, мінез-құлық талдауын және AI (IBM Watson) интеграциясын қатар пайдаланады [7]. Жұмыс принципі мынадай:

1. Логтар QRadar Event/Flow Collectors арқылы жиналады, содан кейін деректер DSM (Device Support Module) көмегімен парсингтен өтеді.

2. Correlation Engine (CRE rules) оқиғалар арасындағы байланысты тауып, соның негізінде инциденттер қалыптастырады. Тереңірек талдау үшін User Behavior Analytics (UBA) қосымша қолданылады.

3. IBM Watson AI тәуекелді бағалап, талдау жүргізеді. Одан кейін инциденттер IBM SOAR-ға жіберіліп, автоматты әрекеттер орындалады.

Жүйенің басты артықшылығы - күрделі корреляция мен мінез-құлық талдауы, ал негізгі шектеуі - жабықтығы және баптауының күрделілігі.

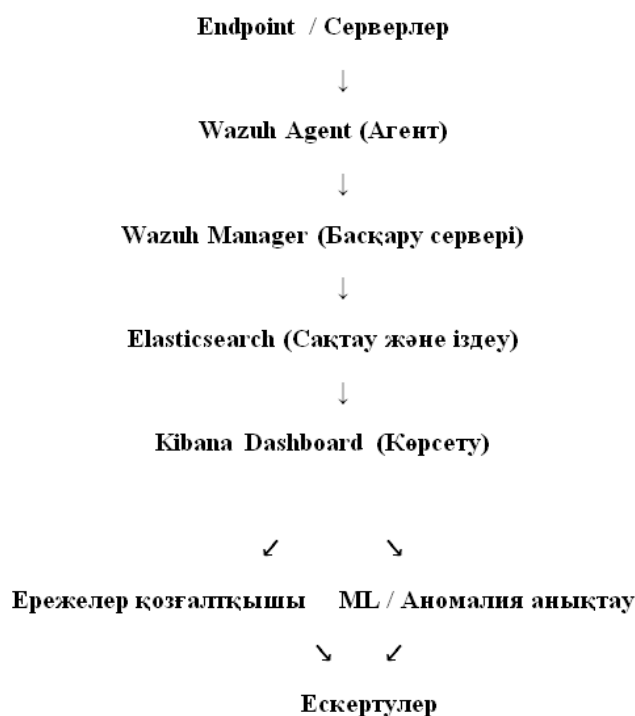


3-сурет – IBM QRadar жүйесі

Ашық бастапқы кодты (open-source) шешімдердің ішінен Wazuh жүйесін бөле айтқан жөн (4-сурет) [11]. Ол Elastic Stack негізінде құрылған, журнал талдауын, интрузияларды анықтауды және машиналық оқыту механизмдерімен интеграцияны қолдайды. Практикалық тұрғыдан бұл шешім ELK базасында іске асыруға ең жақын жүйе. Жұмыс принципі мынадай [8]:

1. Логтар endpoint-тардан Wazuh Agent арқылы жиналып, Wazuh Manager ішінде өңделеді.
2. Rules Engine (XML rules) шабуылдарды анықтайды. Оған қосымша File Integrity Monitoring мен Rootcheck қолданылады.
3. Нәтижелер Kibana/Wazuh Dashboard арқылы ақпараттық қауіпсіздік маманына ұсынылады.

Жүйенің басты артықшылығы - ашық бастапқы коды әрі IDS пен SIEM-нің біріктірілуі, ал негізгі шектеуі - кеңейтілген AI функционалының әзірге жасалмауы [9].

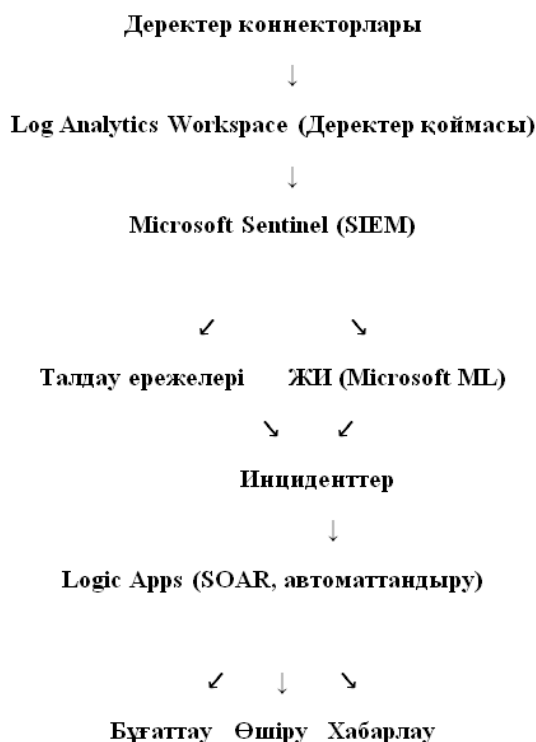


4-сурет – Wazuh лог талдау жүйесі

Бұлтты орта инфрақұрылымдарында Microsoft Sentinel жүйесі белсенді қолданылады, ол SIEM және SOAR функцияларын біріктіреді, сонымен қатар оқиғаларды талдау және жауап беруді автоматтандыру үшін кіріктірілген AI модельдерін қолданады (5-сурет) [15]. Жұмыс принципі келесідей [10]:

1. Деректер Data Connectors (Azure, Office 365, Syslog, API) арқылы жиналады.
2. Логтар Log Analytics Workspace ішінде сақталады, KQL (Kusto Query Language) арқылы талдау жүргізіледі.
3. Analytics Rules инциденттерді анықтайды, UEBA (User Entity Behavior Analytics) қолданушы әрекетін талдайды.
4. Инциденттер SOAR арқылы өңделеді, автоматты әрекеттер (response actions) орындалады.

Ерекшелігі – бұлтты орталарда жиі қолданылады және жақсы масштабталатын шешім, шектеуі – Microsoft экожүйесіне тәуелді [12].



5-сурет – Microsoft Sentinel бұлтты ақпараттық қауіпсіздікті қамтамасыз ету жүйесі

Заманауи зерттеулер, сонымен қатар, үлкен тілдік модельдерге (LLM) негізделген тәсілдердің дамуын көрсетеді. Мысалы, жобада LLaMA сияқты модельдерді қолдану журналдарды талдауға, оқиғаларды түсіндіруге және жауап беру бойынша ұсыныс жасауға мүмкіндік беретіні атап өтіледі [13]. Бұл бағыт бүгінде айрықша өзекті: AI дәл журнал талдауы үшін қолданылады әрі жобада ұсынылған архитектурамен үйлеседі.

Аталған зерттеуде күрделі талдауды қажет етпей-ақ ауытқуларды анықтауға мүмкіндік беретін журнал талдау тәсілі де ұсынылған. Бұл журналды талдау міндеттерінде AI қолданудың тиімділігін тағы бір рет растайды.

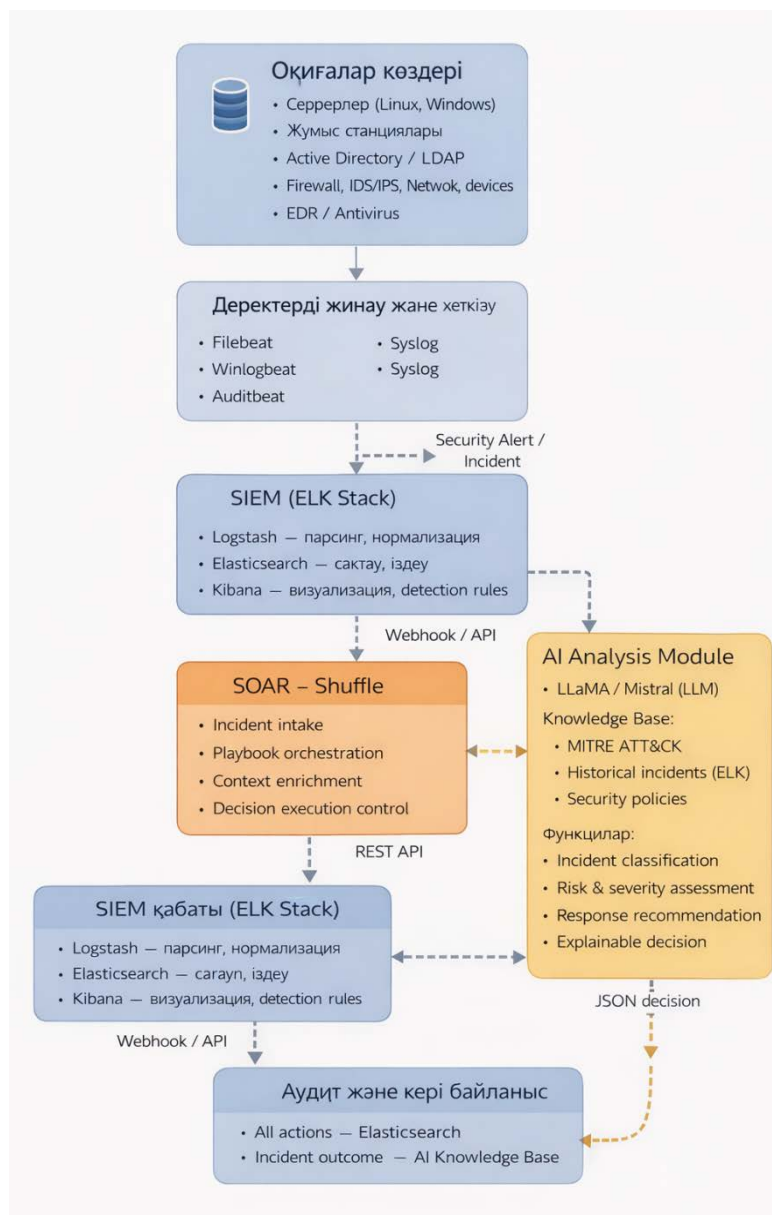
Заманауи жүйелердің тағы бір маңызды буыны - Shuffle секілді SOAR платформалары арқылы жүзеге асатын жауап беруді автоматтандыру [14]. Мұндай платформа автоматты жауап беру сценарийлерін іске қосуға мүмкіндік береді: IP мекенжайларын бұғаттау, пайдаланушыларды өшіру, хабарландыру жіберу және басқа да әрекеттер.

Қорыта айтқанда, қолданыстағы шешімдерді талдау қазіргі ақпараттық қауіпсіздік жүйелерінің SIEM, жасанды интеллект және SOAR-ды қамтитын архитектураға сүйенетінін көрсетеді. Дегенмен көп жағдайда жүйеге дайын модельдер енгізіліп, олар алдын ала анықталған алгоритмдермен шектеледі. Ал арнайы оқытылған модельдерді пайдалану талдау мен автоматтандыру мүмкіндіктерін айтарлықтай кеңейтеді.

Жасалатын жұмыс келесі қадамдардан тұрады (6-сурет):

1. Логтар SIEM жүйесінен (SIEM жүйесі ретінде ELK) алынады.
2. Деректер алдын ала өңдеуден өтеді (parsing, оқиға түріне байланысты жіктеу).
3. SIEM жүйесіндегі ережелер базасы embedding-ке айналып, Qdrant ішіне векторлық мәнге жазылады.
4. Similarity search (деректер арасынан ең ұқсасын табу әдісі) арқылы ұқсас инциденттер табылады.
5. LLM (LLaMA) логтарды семантикалық деңгейде талдайды. Инциденттің мәні, себебі және тәуекелі анықталады. Модель қолданушыға арнап ұсыныстар (recommendations) генерациялайды.
6. Нәтиже SOAR жүйесіне жіберіледі.

Осының нәтижесінде контекстік және семантикалық талдау жүргізіледі, бұл кез келген цифрлық ортаға бейімделгіш және интеллектуалды жүйе ретінде жүзеге асады.



6-сурет – Жоба архитектурасы

Логтар SIEM жүйесінен (ELK) орталықтандырылып жинақталады да, API арқылы жасанды интеллект модуліне жіберіледі. Мұнда SIEM тек дерек көзі рөлін атқарады - осының арқасында қолданыстағы инфрақұрылымды өзгертпей-ақ интеграция жасауға болады әрі жүйе әртүрлі SIEM платформаларымен қатар жұмыс істей алады.

Деректер алдын ала өңделгендіктен, классикалық SIEM-дегідей толық parsing (grok, regex) қажет емес. Жеткілікті болатыны - бірнеше негізгі өріс: уақыт белгісі (timestamp), дерек көзінің IP мекенжайы (source IP) және хабарлама (message). Оның үстіне жүйе raw logs форматымен де жұмыс істей алады, бұл енгізуді жеделдетіп, форматқа тәуелсіздік береді.

Келесі кезеңде логтар embedding форматына түрленеді - әрбір лог жазбасы векторға айналады. Бұл үшін арнайы embedding модельдері қолданылады, нәтижесінде логтар синтаксис бойынша емес, мағынасы бойынша талданады. Осындай тәсіл әртүрлі форматта жазылған ұқсас шабуылдарды тануға және артық шуды сүзуге мүмкіндік береді.

Барлық embedding деректері Qdrant векторлық деректер базасында сақталады. Әр лог үшін embedding пен metadata (IP, уақыт, пайдаланушы) қатар жазылып, индекстеу арқылы жоғары жылдамдықты іздеу қамтамасыз етіледі. Осылайша тарихи контекст жинақталып, жүйені масштабтауға жол ашылады.

Жаңа логтар түскенде similarity search механизмі іске қосылады: жаңа лог embedding-ке айналады, Qdrant ішінен оған ең ұқсас жазбалар табылып, бұрынғы инциденттермен салыстырылады. Дәл осы қадам unknown attack detection мүмкіндігін береді, жалған позитивтерді азайтады және контекстке негізделген талдау жасауға мүмкіндік ашады.

Одан әрі LLM (мысалы, LLaMA) арқылы семантикалық талдау жүргізіледі. Модель логтарды мәтін ретінде оқып, prompt негізінде шабуыл түрін, қауіптілік деңгейін және оның себептерін анықтайды. Талдау сапасын арттыру үшін модельге контекст ретінде ұқсас логтар қоса беріледі - бұл нәтижені түсіндірмелі түрде ұсынуға көмектеседі.

Нәтижесінде жүйе инцидент түрін (мысалы, Brute Force немесе SQL Injection), негізгі себебін (root cause) және тәуекел деңгейін (low, medium, high) айқындайды. Мұндай құрылымдалған әрі түсінікті ақпарат SOC аналитиктеріне инциденттерді басымдыққа қарай сұрыптауды жеңілдетеді.

Осы деректерге сүйене отырып, жасанды интеллект ұсыныстарды автоматты қалыптастырады - IP мекенжайды бұғаттау, пайдаланушыны өшіру немесе қосымша тексеру жүргізу сияқты әрекеттерді нұсқайды. Бұл аналитиктердің жүктемесін жеңілдетеді, әрекет ету уақытын қысқартады және стандартты процестерді автоматтандыруға мүмкіндік береді.

Соңғы кезеңде барлық нәтиже SOAR жүйесіне (мысалы, Shuffle) жіберіледі. Онда алдын ала анықталған playbook-тар іске қосылып, автоматты әрекеттер орындалады. Соның арқасында процесс толық автоматтандырылып, қауіпке әрекет ету уақыты айтарлықтай қысқарады әрі инциденттер тез оқшауланады.

ELK, SOAR және AI негізіндегі орталықтандырылған басқару архитектурасы қазіргі киберқауіптерге қарсы күрестегі күрделі мәселелерге кешенді, тиімді шешім ұсынады. Мұндай тәсіл ұйымдарға қауіптерді дер кезінде анықтауға, оларға жауап беруге және ақпараттық активтерін қорғауға мүмкіндік береді.

Қорытынды

Қорыта айтқанда, жасанды интеллектке негізделген SIEM жүйелері қауіптерді алдын ала болжай алатындықтан, дәстүрлі құралдармен салыстырғанда тәуекелді әлдеқайда төмендетеді. Мұндай интеллектуалды жүйелер қауіпті ерте кезеңде анықтауда, жауап беруді жеделдетуде және операциялық тиімділікті автоматты арттыруда өзін тиімді құрал ретінде көрсетті. Олар үлкен көлемді деректерді талдап, ауытқуларды тауып, көп еңбекті талап ететін біртекті міндеттерді автоматтандырады - осылайша қауіпсіздік мамандарын стратегиялық шешім қабылдауға босатады. Зиянды бағдарламаларды талдаудан бастап осалдықтарды басқаруға дейінгі кең ауқымды жұмысты осындай жүйелер атқара алады.

Сонымен қатар AI-мен жұмыс істейтін SIEM жүйелері дамып келе жатқан қауіптерге қарсы кешенді қорғаныс қызметін атқарады. Оларды енгізу арқылы ұйымдар үнемі өзгеріп отыратын киберландшафтта сенімділігін нығайтып, күрделі қатерлерді дер кезінде аулай алады әрі тәуекелге төзімділігін арттырады.

Әдебиеттер тізімі

1. Vardalachakis M., Vasilakis M., Tampouratzis M. A Comprehensive Analysis of Features, Benefits, Challenges, and Best Practices of Security Information and Event Management (SIEM) Solutions // Computer Sciences & Mathematics Forum. – 2025. – Vol. 12, № 1. – Art. 18. – <https://doi.org/10.3390/cmsf2025012018>

2. Singh S. et al. A future-proof approach to cybersecurity compliance: The power of AI and ML in SIEM, SOAR and cloud SOC. – ResearchGate, 2024. – P. 2–6.
3. Pan J. AI based Log Analyser: A Practical Approach // arXiv preprint arXiv:2203.10960. – 2022. – <https://doi.org/10.48550/arXiv.2203.10960>
4. Ramakrishnan S., Chittibala D. R. Enhancing Cyber Resilience: Convergence of SIEM, SOAR, and AI in 2024 // International Journal of Computing and Engineering. – 2024. – Vol. 5, № 2. – P. 36–44. – <https://doi.org/10.47941/ijce.1754>
5. da Silva G. R., de Oliveira B. M. C., Braga R. Comparison of Machine Learning and the ELK Stack for Automated Cyber Attack Detection: A Comparative Approach // 12th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2025), Nice, France. – 2025. – P. 14–20. – https://doi.org/10.48545/advance2025-fullpapers-1_2.
6. Sahay R., Sumasadan M., Eapen B., Meng W., Al Mamu M. R. Enhancing Threat Hunting with Splunk and Generative AI for Automated Security Operations // Research Square (preprint). – 2025. – <https://doi.org/10.21203/rs.3.rs-7515771/v1>
7. Alliouche R., Chenni H. Toward Intelligent Cyber Defense: A Comprehensive Study of SOAR Technologies for AI-Based DDoS Detection. – ResearchGate, 2025. – P. 12–13..
8. Patel D. G., Pujari S. R. AI-driven incident response in cloud security // International Journal of Science and Research Archive. – 2025. – Vol. 15, № 3. – P. 1463–1475. – <https://doi.org/10.30574/ijrsra.2025.15.3.1742>
9. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. – 2009. – Vol. 41, № 3. – Art. 15. – <https://doi.org/10.1145/1541880.1541882>
10. Dakić V., Morić Z., Kapulica A., Regvart D. Leveraging Microsoft Sentinel and Logic Apps for Automated Cyber Threat Response // Edelweiss Applied Science and Technology. – 2024. – Vol. 8, № 6. – P. 4319–4348. – <https://doi.org/10.55214/25768484.v8i6.2933>
11. Habibzadeh A., Rahnema-Moghadam S., Beheshti A. Large Language Models for Security Operations Centers: A Comprehensive Survey // arXiv preprint. – 2025. – arXiv:2509.10858. – <https://doi.org/10.48550/arXiv.2509.10858>
12. Loumachi F. Y., Ghanem M. C., Ferrag M. A. Advancing Cyber Incident Timeline Analysis Through Retrieval-Augmented Generation and Large Language Models // Computers. – 2025. – Vol. 14, № 2. – Art. 67. – P. 1–42. – <https://doi.org/10.3390/computers14020067>
13. Guan W., Wang Y., Zhang X. LogLLM: Log-based Anomaly Detection Using Large Language Models // arXiv preprint. – 2024. – arXiv:2411.08561. – <https://doi.org/10.48550/arXiv.2411.08561>
14. Saha D., Tarek S., Yahyaei K., Saha S. K., Zhou J., Tehranipoor M., Farahmandi F. LLM for SoC Security: A Paradigm Shift // IEEE Access. – 2024. – Vol. 12. – P. 155498–155521. – <https://doi.org/10.1109/ACCESS.2024.3427369>
15. Tendikov N., Rzayeva L., Saoud B., Shaye A., Azmi M. H., Myrzatay A., Alnakhli M. Security Information Event Management Data Acquisition and Analysis Methods with Machine Learning Principles // Results in Engineering. – 2024. – Vol. 22. – Art. 102254. – <https://doi.org/10.1016/j.rineng.2024.102254>

A.Kushubek, B.Sharipbekov, B.Rakhimov, S. Santeeva, A.Shaihanova

A REVIEW OF ARTIFICIAL INTELLIGENCE-BASED LOG ANALYSIS AND AUTOMATED RESPONSE ARCHITECTURES IN IT INFRASTRUCTURE INFORMATION SECURITY MANAGEMENT

The cybersecurity industry has undergone significant changes over the past decade: along with the rapid development of technologies, the number of new threats is also increasing. Organizations are simultaneously faced with insider threats, online espionage, malware, and data leakage. Since the activities of an enterprise rely on interconnected information systems, a cyberattack can disrupt the entire workflow and cause significant financial losses. In this regard, companies are increasingly paying attention to preventive measures to protect their confidential data and digital property. The article considers methods and architecture for centralized management of information security of IT infrastructure using artificial intelligence technologies. The main focus is on the use of SIEM systems to collect and correlate security events, as well as the integration of intelligent algorithms for detecting anomalies and analyzing events. An architectural project is proposed that combines an event management system, an artificial intelligence

module, and a SOAR platform for automated response. Practical scenarios for detecting and mitigating threats based on log analysis, as well as the possibilities of automating protective measures, are analyzed. The results show that intelligent analytics and process automation can increase the efficiency of incident detection, reduce the workload of information security professionals, and shorten response times.

Keywords: attacks, information security, SIEM, artificial intelligence, SOAR, automation, security events.

А.Кушубек, Б.Шарипбеков, Б.Рахимов, С. Сантеева, А.Шайханова

ОБЗОР АРХИТЕКТУРЫ ЛОГ-АНАЛИЗА И АВТОМАТИЗИРОВАННОГО РЕАГИРОВАНИЯ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В УПРАВЛЕНИИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТЬЮ ИТ-ИНФРАСТРУКТУР

За последнее десятилетие индустрия кибербезопасности претерпела значительные изменения: наряду с быстрым развитием технологий растет и число новых угроз. Организации одновременно сталкиваются с внутренними угрозами, онлайн-шпионажем, вредоносным ПО и утечками данных. Поскольку деятельность предприятия зависит от взаимосвязанных информационных систем, кибератака может нарушить весь рабочий процесс и привести к значительным финансовым потерям. В связи с этим компании все больше внимания уделяют превентивным мерам по защите своих конфиденциальных данных и цифровой собственности. В статье рассматриваются методы и архитектура централизованного управления информационной безопасностью ИТ-инфраструктуры с использованием технологий искусственного интеллекта. Основное внимание уделяется использованию SIEM-систем для сбора и сопоставления событий безопасности, а также интеграции интеллектуальных алгоритмов для обнаружения аномалий и анализа событий. Предлагается архитектурный проект, объединяющий систему управления событиями, модуль искусственного интеллекта и платформу SOAR для автоматизированного реагирования. Анализируются практические сценарии обнаружения и смягчения угроз на основе анализа журналов, а также возможности автоматизации защитных мер. Результаты показывают, что интеллектуальная аналитика и автоматизация процессов могут повысить эффективность обнаружения инцидентов, снизить рабочую нагрузку специалистов по информационной безопасности и сократить время реагирования.

Ключевые слова: атаки, информационная безопасность, SIEM, искусственный интеллект, SOAR, автоматизация, события безопасности.

References

1. Alliouche, R., & Chenni, H. (2025). Toward Intelligent Cyber Defense: A Comprehensive Study of SOAR Technologies for AI-Based DDoS Detection. ResearchGate, 12-13.
2. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), Article 15. <https://doi.org/10.1145/1541880.1541882>
3. da Silva, G. R., de Oliveira, B. M. C., & Braga, R. (2025). Comparison of Machine Learning and the ELK Stack for Automated Cyber Attack Detection: A Comparative Approach. 12th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2025), Nice, France, 14–20. https://doi.org/10.48545/advance2025-fullpapers-1_2
4. Dakić, V., Morić, Z., Kapulica, A., & Regvart, D. (2024). Leveraging Microsoft Sentinel and Logic Apps for Automated Cyber Threat Response. Edelweiss Applied Science and Technology, 8(6), 4319–4348. <https://doi.org/10.55214/25768484.v8i6.2933>
5. Guan, W., Wang, Y., & Zhang, X. (2024). LogLLM: Log-based Anomaly Detection Using Large Language Models (arXiv:2411.08561). arXiv. <https://doi.org/10.48550/arXiv.2411.08561>
6. Habibzadeh, A., Rahnama-Moghadam, S., & Beheshti, A. (2025). Large Language Models for Security Operations Centers: A Comprehensive Survey (arXiv:2509.10858). arXiv. <https://doi.org/10.48550/arXiv.2509.10858>

7. Loumachi, F. Y., Ghanem, M. C., & Ferrag, M. A. (2025). Advancing Cyber Incident Timeline Analysis Through Retrieval-Augmented Generation and Large Language Models. *Computers*, 14(2), Article 67, 1–42. <https://doi.org/10.3390/computers14020067>
8. Pan, J. (2022). AI based Log Analyser: A Practical Approach (arXiv:2203.10960). *arXiv*. <https://doi.org/10.48550/arXiv.2203.10960>
9. Patel, D. G., & Pujari, S. R. (2025). AI-driven incident response in cloud security. *International Journal of Science and Research Archive*, 15(3), 1463–1475. <https://doi.org/10.30574/ijrsra.2025.15.3.1742>
10. Ramakrishnan, S., & Chittibala, D. R. (2024). Enhancing Cyber Resilience: Convergence of SIEM, SOAR, and AI in 2024. *International Journal of Computing and Engineering*, 5(2), 36–44. <https://doi.org/10.47941/ijce.1754>
11. Saha, D., Tarek, S., Yahyaei, K., Saha, S. K., Zhou, J., Tehranipoor, M., & Farahmandi, F. (2024). LLM for SoC Security: A Paradigm Shift. *IEEE Access*, 12, 155498–155521. <https://doi.org/10.1109/ACCESS.2024.3427369>
12. Sahay, R., Sumasadan, M., Eapen, B., Meng, W., & Al Mamu, M. R. (2025). Enhancing Threat Hunting with Splunk and Generative AI for Automated Security Operations (Preprint). *Research Square*. <https://doi.org/10.21203/rs.3.rs-7515771/v1>
13. Singh, S., et al. (2024). A future-proof approach to cybersecurity compliance: The power of AI and ML in SIEM, SOAR and cloud SOC. *ResearchGate*, 2–6.
14. Tendikov, N., Rzaeva, L., Saoud, B., Shaya, I., Azmi, M. H., Myrzatay, A., & Alnakhli, M. (2024). Security Information Event Management Data Acquisition and Analysis Methods with Machine Learning Principles. *Results in Engineering*, 22, Article 102254. <https://doi.org/10.1016/j.rineng.2024.102254>
15. Vardalachakis, M., Vasilakis, M., & Tampouratzis, M. (2025). A Comprehensive Analysis of Features, Benefits, Challenges, and Best Practices of Security Information and Event Management (SIEM) Solutions. *Computer Sciences & Mathematics Forum*, 12(1), Article 18. <https://doi.org/10.3390/cmsf2025012018>

Авторлар туралы мәлімет:

Кушубек Аслан Рашидулы (автор-корреспондент) – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, бас сарапшы, Астана, Қазақстан, e-mail: kushubek.aslan@mail.ru

Шарипбеков Беркин Ерсейтулы – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, бөлім басшысы, Астана, Қазақстан, e-mail: b.sharipbekov@testcenter.kz

Рахимов Батырбек Серикович – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, директор орынбасары, Астана, Қазақстан, e-mail: rakhimov.batyrbek@mail.ru

Сантеева Сая Әділбайқызы – Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Ақпараттық қауіпсіздік кафедрасының аға оқытушысы, Астана, Қазақстан, e-mail: saya_santeeva@mail.ru

Шайханова Айгүл Кайрулаевна – Л.Н. Гумилев атындағы Еуразия ұлттық университеті, PhD, қауымдастырылған профессор, Ақпараттық қауіпсіздік кафедрасының профессоры, Астана, Қазақстан, e-mail: aigul.shaikhanova@gmail.com

Information about authors:

Kushubek Aslan Rashiduly (corresponding author) – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Chief expert, Astana, Kazakhstan, e-mail: kushubek.aslan@mail.ru

Sharipbekov Berkin Erseituly – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Department Head, e-mail: b.sharipbekov@testcenter.kz

Rakhimov Batyrbek Serikovich – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Deputy Director, e-mail: rakhimov.batyrbek@mail.ru

Santeeva Saya – L.N. Gumilyov Eurasian National University, Senior Lecturer, Department of Information Security, Astana, Republic of Kazakhstan, e-mail: saya_santeeva@mail.ru

Shaikhanova Aigul – L.N. Gumilyov Eurasian National University, PhD, Associate Professor, Professor of the Department of Information Security, Astana, Republic of Kazakhstan, e-mail: aigul.shaikhanova@gmail.com

Сведения об авторах:

Кушубек Аслан Рашидулы (автор-корреспондент) – РГП на ПХВ «Национальный центр тестирования» МНВО РК, главный эксперт, Астана, Казахстан, e-mail: kushubek.aslan@mail.ru

Шарипбеков Беркин Ерсейтулы – РГП на ПХВ «Национальный центр тестирования» МНВО РК, руководитель отдела, Астана, Казахстан e-mail: b.sharipbekov@testcenter.kz

Рахимов Батырбек Серикович – РГП на ПХВ «Национальный центр тестирования» МНВО РК, заместитель директора, Астана, Казахстан, e-mail: rakhimov.batyrbek@mail.ru

Сантеева Сая Әділбайқызы – ЕНУ им. Л. Н. Гумилева, старший преподаватель кафедры информационной безопасности, Астана, Казахстан, e-mail: saya_santeeva@mail.ru

Шайханова Айгуль Кайрулаевна – ЕНУ им. Л. Н. Гумилева, PhD, ассоц. профессор, профессор кафедры информационной безопасности, Астана, Казахстан, e-mail: aigul.shaikhanova@gmail.com

A. Zhorabekova¹, M. Duisenova^{2*}, M. Kurmasheva³, A. Yesnazar⁴

¹M. Auezov South Kazakhstan University, Shymkent, Kazakhstan

^{2,3,4}O. Zhanibekov South Kazakhstan Pedagogical University, Shymkent, Kazakhstan

*e-mail: duisenova.marzhan@okmpu.kz

¹ORCID 0000-0003-2697-8077, ²ORCID 0000-0001-5684-7364

³ORCID 0000-0002-4965-8490, ⁴ORCID 0000-0002-8658-0135

MEASUREMENT INVARIANCE OF AN AI-GAMIFIED ADAPTIVE LEARNING ACCEPTANCE SCALE ACROSS ENGLISH PROFICIENCY AND YEAR OF STUDY

Students' acceptance of adaptive learning tools that combine artificial intelligence with gamification is usually measured with multidimensional questionnaires, but whether such questionnaires show the same measurement properties across student subgroups is rarely examined. This study tested the measurement invariance of a 13-item scale of acceptance of AI-gamified adaptive learning across English proficiency and year of study and estimated latent mean differences where invariance held. Previously collected data from 401 EFL students at a Kazakhstani pedagogical university were analysed with multi-group confirmatory factor analysis. A second-order factor model served as the primary measurement model, and a four-correlated-factor model estimated with MLR and WLSMV was used as a sensitivity analysis. Across proficiency levels A2, B1 and B2+, the scale showed full scalar invariance, with CFI decreasing by no more than .003 and no significant chi-square difference at any step, and the latent mean of general acceptance did not differ between levels. Across year of study, comparing first-year students with those in years two to four, metric invariance was supported and scalar invariance was partial. Once the intercept of one item concerning AI feedback was freed, constrained and unconstrained models no longer differed significantly. Within the reported partial invariance model, the latent mean of general acceptance was higher for upper-year students than for first-year students by about 0.3 standard deviations. The findings provide a full psychometric basis for comparisons across proficiency and a partial basis for latent mean comparisons across year of study. The main limitations are a single-institution convenience sample, self-reported proficiency, a ceiling effect and the possibility of common-method variance.

Keywords: measurement invariance, multi-group confirmatory factor analysis, second-order factor model, technology acceptance, artificial intelligence, gamification, adaptive learning, foreign language teaching, psychometrics.

Introduction

Foreign language teaching has seen the rapid spread of digital tools in which adaptive, AI-driven sequencing is paired with game mechanics such as points, levels and rewards. The pedagogical value of such a tool is likely to depend on how far students accept it, since acceptance affects the probability of sustained use. Measuring acceptance has therefore become one of the main instruments in research on AI-gamified adaptive learning. The technology acceptance models TAM [1] and UTAUT [2] treat acceptance as a construct with several components. The gamification literature [3] singles out game elements as a separately perceived aspect. Questionnaires used in EFL contexts tend to follow this tradition and are usually designed as multidimensional.

Evidence on the quality of such questionnaires, however, is mostly limited to internal consistency coefficients and to the factor structure obtained in a single sample. The standards for educational and psychological testing [4] require evidence that an instrument works in the same way in the groups whose scores are to be compared. In psychometrics this requirement is expressed through the concept of measurement invariance, understood as the independence of the relation between observed item responses and the latent construct from group membership [5, 6]. When invariance fails, a mean difference between groups may reflect a different reading of the items rather than a real difference in the construct. The justification for interpreting any between-group

comparison, for instance between students of different proficiency or study experience, is therefore limited until invariance has been tested [7].

The established approach to testing invariance is multi-group confirmatory factor analysis, in which constraints are imposed on model parameters step by step. Configural invariance means an identical factor structure, metric invariance means equal factor loadings, and scalar invariance means equal item intercepts [5, 8]. Scalar invariance is required for comparing latent means [6]. When full invariance is not supported, partial invariance is applied. At least two items per factor are kept invariant while the remaining constraints are released [9]. For second-order models the testing sequence is longer and covers first-order and second-order loadings as well as intercepts at the item and factor levels [10]. Alongside the chi-square difference test, criteria that are less sensitive to sample size are used to compare nested models. The most common are $\Delta CFI \leq .01$ [11] and $\Delta RMSEA \leq .015$ [12].

The study that this article builds on [13] examined the measurement structure of a 13-item scale of acceptance of AI-gamified adaptive learning, collected from 401 EFL students at a Kazakhstani pedagogical university. Model comparison showed that the four hypothesised subdomains could not be separated empirically. These subdomains covered acceptance of gamification, acceptance of AI, effectiveness, and engagement together with satisfaction. Latent correlations reached .95, and in the bifactor model the general factor was dominant, with ECV at .82 and ω at .93. On this basis the scale was recommended to be scored as a single general score. The same study openly listed the absence of invariance testing as a limitation and formulated it as a task for future work. This gap matters. If the scale is used as a total score, comparing students who differ in English proficiency or year of study is justified only when the equivalence of item loadings and intercepts across these groups has been confirmed. Invariance across proficiency is especially relevant here. The items describe AI and adaptive tools, and students at lower and higher levels may interpret these descriptions through different experience.

This article applies an analysis not reported before to the data of [13], multi-group confirmatory factor analysis. The second-order factor model was chosen as the primary measurement model because the earlier study recommended scoring the scale as a general score, and a second-order model allows the latent mean of the general acceptance factor to be compared across groups. The four-correlated-factor model serves as a sensitivity analysis. The grouping variables are English proficiency and year of study. The **aim of the study** is to test the measurement invariance of the AI-gamified adaptive learning acceptance scale across these groups and, where invariance is supported, to estimate latent mean differences. The **research questions** are as follows.

RQ1. Does the second-order acceptance model fit separately in each proficiency group A2, B1 and B2+, that is, does configural invariance hold?

RQ2. Are item loadings, second-order loadings and intercepts equivalent across these levels? If full invariance is not supported, which items violate partial invariance?

RQ3. Given at least partial scalar invariance, does the latent mean of general acceptance differ between levels?

RQ4. Does the same testing sequence replicate across year of study, that is, between first-year and upper-year students?

Materials and methods

The study is a secondary psychometric analysis of previously collected cross-sectional survey data. The data come from 401 EFL students of O. Zhanibekov South Kazakhstan Pedagogical University, surveyed via Google Forms. In the questionnaire each statement appeared in three languages at once, English together with Kazakh and Russian, so that respondents read a single trilingual item rather than a language-specific version. Participation was voluntary and responses were stored anonymously. Of the initial 421 questionnaires, 20 were removed because consent was not given or item responses were missing. The factor structure and measurement quality of these data have been published before [13]. That publication did not test measurement invariance and

listed it as a task for future work. The present article applies an analysis to the same data that has not been reported anywhere. The original structural results, that is, model comparison, discriminant validity and bifactor indices, are not repeated here and are only cited.

Of the participants, 90.3 percent were female ($n = 362$) and 70.1 percent were first-year students ($n = 281$). English proficiency was collected through self-assessment. Twenty-eight respondents who marked two levels at once were assigned to the higher level. This coding decision was made within the present study and reproduces exactly the group counts reported in Table 1 of [13]. Fourteen respondents who reported C1 or above were merged into the B2 group to keep the group size workable. The first grouping thus consists of A2 ($n = 129$), B1 ($n = 151$) and B2+ ($n = 121$). The second grouping compares first-year students ($n = 281$) with students of years two to four ($n = 120$), the latter group containing 88 second-year, 23 third-year and 9 fourth-year students. Grouping by gender was not possible because of the small number of male students ($n = 35$). The two grouping variables are related. Of the A2 students, 110 study in the first year, while in the B2+ group the shares of first-year and upper-year students are comparable, at 68 and 53. This circumstance is taken into account in the discussion.

The 13-item scale covers four hypothesised domains. Acceptance of gamification is measured by items GAM1 to GAM3, acceptance of AI by items AI1 to AI3, effectiveness by items EFF1 to EFF3, and engagement together with satisfaction by items ENG1 to ENG3 and SAT. The items refer to the general class of AI-gamified adaptive learning tools that students had encountered in their studies rather than to any named platform. The scale therefore measures general acceptance of this type of tool, not the evaluation of a particular product. Responses were given on a five-point Likert scale, where 1 means complete disagreement and 5 means complete agreement. The full item wording and the domain assignment are given in [13]. In the earlier analysis item means ranged from 4.03 to 4.31, distributions were negatively skewed (-0.77 to -1.26), and overall $\alpha = .958$. In the present study the internal consistency of the 13 items was computed separately for each group.

The primary model is the second-order factor model of [13], in which the four first-order factors GAM, AI, EFF and ENG load on one general acceptance factor G. This model was chosen for two reasons. The earlier study recommended scoring the scale as one general score, and a second-order model makes it possible to compare the latent mean of the general factor across groups [10]. The bifactor model was judged unsuitable for invariance testing because several specific loadings in [13] fell between .06 and .16 and the model was not identified under WLSMV. The one-factor model showed poor fit there ($RMSEA = .137$). The four-correlated-factor model was used as a sensitivity analysis.

For the second-order model the sequence of Chen, Sousa and West was followed [10]. In the configural model the structure is identical and all parameters are estimated freely in each group. The second step equates first-order loadings, and the third adds the equality of second-order loadings. The fourth step equates item intercepts, while the intercepts of the first-order factors remain free in the second group and the mean of G stays fixed at zero in all groups. In the fifth step the intercepts of the first-order factors are fixed at zero in all groups, and the latent mean of G is fixed at zero in the reference group and estimated freely in the others. All partial variants were estimated within this same identification, by removing the relevant intercept equality. The first-order factors and G were identified by the marker method through their first indicators. In the configural and metric models all latent means were fixed at zero. For the four-factor model the standard configural, metric and scalar sequence was used [5, 8]. When full scalar invariance was not supported, partial invariance was tested [9]. The intercept with the largest contribution according to the score test was identified and released.

Maximum likelihood estimation robust to non-normality, MLR, served as the main estimator. Scaled χ^2 , CFI, TLI, RMSEA with a 90 percent confidence interval and SRMR are reported [14]. Nested models were compared with the Satorra and Bentler corrected χ^2 difference test [15] and with criteria that are less sensitive to sample size, $\Delta CFI \geq -.01$ [11] and $\Delta RMSEA \leq .015$ [12]. Where the criteria disagree, both are reported. To account for the ordinal nature of the items, the four-factor model was re-estimated with WLSMV under the identification of Wu and Estabrook

[16]. In that sequence the configural model is followed by equal thresholds, then equal loadings and finally equal intercepts. Because category 2 of item GAM2 was empty in the B2+ group, categories 1 and 2 of all items were merged in this sensitivity analysis. The analysis was run in R with the packages lavaan 0.6.17 [17] and semTools 0.5.6 [18]. The data contain no missing values.

Results and discussion

Internal consistency of the 13 items was high in every group. In the A2 group $\alpha = .966$, in B1 $\alpha = .955$, in B2+ $\alpha = .951$, among first-year students $\alpha = .957$ and among upper-year students $\alpha = .959$. The overall mean score of the 13 items was practically identical across levels, at 4.18 for A2, 4.18 for B1 and 4.20 for B2+. Across year of study it was slightly higher among upper-year students, at 4.33 against 4.13.

The second-order model fitted satisfactorily in the three groups simultaneously (Table 1). As the constraints were added step by step, CFI decreased by at most .003 and RMSEA by .002. The χ^2 difference test was not significant at any step. First-order and second-order loadings, item intercepts and the intercepts of the first-order factors proved equivalent across A2, B1 and B2+, which means that full scalar invariance was supported (RQ1, RQ2). Relative to A2, the latent mean of the general acceptance factor was +0.004 in B1 (SE = 0.072; $p = .96$) and +0.004 in B2+ (SE = 0.076; $p = .96$). No difference between levels was observed (RQ3).

Table 1 - Invariance of the second-order model across English proficiency, groups A2, B1 and B2+, MLR, scaled indices

Model	χ^2 (df)	CFI	TLI	RMSEA [90% CI]	SRMR	Δ CFI	Δ RMSEA	$\Delta\chi^2$ (Δ df), p
1. Configural	296.7 (183)	0.953	0.939	0.068 [0.057; 0.079]	0.041	-	-	-
2. + first-order loadings	318.9 (201)	0.951	0.943	0.066 [0.055; 0.077]	0.054	-0.002	-0.002	19.0 (18), .39
3. + second-order loadings	325.5 (207)	0.951	0.944	0.065 [0.055; 0.076]	0.066	0.000	-0.001	6.6 (6), .36
4. + item intercepts	349.1 (225)	0.948	0.946	0.064 [0.054; 0.074]	0.067	-0.003	-0.001	20.5 (18), .31
5. + first-order factor intercepts	356.9 (231)	0.948	0.947	0.064 [0.053; 0.074]	0.068	0.000	0.000	6.8 (6), .34
Note. $\Delta\chi^2$ is the Satorra and Bentler corrected difference test relative to the preceding model								

The four-factor model repeated this conclusion. In the configural model CFI = .953 and RMSEA = .069. In the metric model Δ CFI = -.002, $\Delta\chi^2(18) = 18.3$, $p = .43$. In the scalar model Δ CFI = -.002, $\Delta\chi^2(18) = 20.5$, $p = .31$. None of the latent means of the four first-order factors differed between groups ($|est| \leq 0.10$; $p \geq .26$). In the WLSMV sensitivity analysis the CFI of the models with equal thresholds, loadings and intercepts stayed stable between .990 and .992, with Δ CFI $\leq .001$. RMSEA fell from .100 to .074 as the constraints were added.

The pattern was more complex for year of study (Table 2). The equality of first-order loadings was supported by the Δ CFI (-.003) and Δ RMSEA (.000) criteria, but the χ^2 difference was significant at the margin ($p = .048$). The equality of second-order loadings was fully supported. The equality of item intercepts was supported by Δ CFI = -.004 and Δ RMSEA = +.002, yet $\Delta\chi^2(9) = 23.5$, $p = .005$. The score test located the largest violation in the intercept of item AI3 ($X^2 = 7.6$; $p = .006$), with item AI1 in second place ($X^2 = 4.7$). Partial scalar model 4b, in which only the AI3 intercept was freed, did not differ significantly from the metric model ($\Delta\chi^2(8) = 14.9$; $p = .062$; CFI = .960). Two final models were used to estimate the latent mean. In model 5a, built

on the full scalar basis, the latent mean of general acceptance was higher for upper-year students than for first-year students by +0.158 (SE = 0.064; $p = .014$), which is about 0.28 standard deviations. In partial model 5b, with the AI3 intercept left free, the difference was +0.155 (SE = 0.064; $p = .016$), about 0.27 standard deviations. The two estimates are practically identical. One nuance deserves mention. In the partial branch, the step from model 4b to model 5b, that is, expressing the group difference through the general factor alone, was violated at the margin by $\Delta\chi^2$ ($\Delta\chi^2(3) = 10.9$; $p = .012$) while staying within $\Delta CFI = -.003$. This agrees with the four-factor result reported below, where the year difference is absent in the AI factor.

Table 2 - Invariance of the second-order model across year of study, first year and years 2-4, MLR, scaled indices

Model	χ^2 (df)	CFI	TLI	RMSEA [90% CI]	SRMR	ΔCFI	$\Delta RMSEA$	$\Delta\chi^2$ (Δdf), p
1. Configural	200.5 (122)	0.965	0.956	0.057 [0.046; 0.067]	0.035	-	-	-
2. + first-order loadings	216.9 (131)	0.962	0.955	0.057 [0.047; 0.067]	0.047	-0.003	0.000	17.1 (9), 0.048
3. + second-order loadings	219.4 (134)	0.962	0.956	0.056 [0.046; 0.066]	0.055	0.000	-0.001	2.4 (3), 0.50
4a. + item intercepts (full)	239.0 (143)	0.958	0.954	0.058 [0.048; 0.067]	0.056	-0.004	+0.002	23.5 (9), 0.005
4b. + item intercepts (partial, AI3 free)	233.5 (142)	0.960	0.956	0.057 [0.047; 0.066]	0.055	-0.002	+0.001	14.9 (8), 0.062
5a. 4a + first-order factor intercepts	243.8 (146)	0.957	0.954	0.058 [0.048; 0.067]	0.057	-0.001	0.000	4.6 (3), 0.20
5b. 4b + first-order factor intercepts	241.6 (145)	0.957	0.954	0.058 [0.048; 0.067]	0.057	-0.003	+0.001	10.9 (3), 0.012

Note. Models 4a and 4b are compared with model 3. Model 5a is compared with 4a, and model 5b with 4b. All partial models were estimated within the identification of the main sequence

The four-factor model repeated this pattern. In the metric model $\Delta\chi^2(9) = 16.7$, $p = .054$, $\Delta CFI = -.003$. In the scalar model $\Delta\chi^2(9) = 23.5$, $p = .005$, $\Delta CFI = -.005$. In the partial scalar model with the AI3 intercept freed, $\Delta\chi^2(8) = 14.8$, $p = .06$, CFI = .960. The latent means of the first-order factors were higher among upper-year students for GAM (+0.17; $p = .034$), EFF (+0.26; $p = .008$) and ENG (+0.25; $p = .010$). No difference was found for AI (+0.06; $p = .51$). In the WLSMV sensitivity analysis CFI stayed stable at .992 to .993, with $\Delta CFI \leq .001$.

In the second-order model the disturbance variance of the EFF factor was estimated close to zero in several groups and slightly negative in some. In the B2+ group it was -0.012 (SE = 0.030), and among first-year students -0.003 (SE = 0.024). These estimates do not differ from zero statistically. Such a result agrees with the conclusion of [13], where an S-1 bifactor diagnostic implicated the effectiveness-specific factor as the weakest component and the specific loadings of the effectiveness items dropped to .06. The effectiveness factor therefore carries very little variance of its own beyond general acceptance. Fixing a boundary estimate at zero is a standard remedy [19]. In the sensitivity analysis with the EFF disturbance fixed at zero, every conclusion stayed unchanged, and the G difference across year of study was +0.126 ($p = .022$).

The results support two conclusions. The scale proved fully invariant with respect to English proficiency. Since the equivalence of item loadings and intercepts was supported, latent means can be compared directly across levels. That comparison found no difference. This agrees with the non-significant proficiency coefficient in the regression of [13] ($p = .80$) and supports, at the latent level, the view that acceptance depends on a general evaluative orientation rather than on language preparation. Invariance with respect to year of study is partial. The intercept of item AI3, the statement that AI feedback helps to correct mistakes more effectively, differs in the upper years. The reason for this difference is hard to state with confidence. One possible explanation is that upper-year students have more experience of receiving feedback from AI tools. For them the item may carry experiential content, whereas first-year students may read it as a general attitude statement. This conjecture cannot be tested with the present data. In technology acceptance research it is established practice, under partial scalar invariance, to compare groups in exactly this way, through latent means in a model with the non-invariant item freed [9, 20]. The conclusions about year of study were therefore drawn for latent means. When an observed total score is compared across years directly, the contribution of item AI3 should be treated with caution.

The advantage of about 0.3 standard deviations in acceptance among upper-year students points in the same direction as the marginal positive year coefficient in the regression of [13] ($p = .060$). There is little ground, however, for reading this difference as an effect of year of study itself. Proficiency and year are related in the sample, as 85 percent of A2 students are in the first year. The upper-year group is small and internally heterogeneous. Frequency of use of AI tools was the only significant predictor of acceptance in [13], and it may change together with year of study. The four-factor model also showed that the difference is uneven across domains. It is present in the effectiveness, engagement and gamification factors and absent in the AI factor. The latent mean difference is therefore better read as evidence that comparison across years has a psychometric basis than as a causal claim that acceptance grows with year of study.

The study rests on a convenience sample from one university, composed mostly of female and first-year students. The year groups are unequal, at 281 and 120, which may make the estimates in the smaller group less precise. English proficiency was determined by self-assessment rather than a standardised test, and the small group reporting C1 or above was merged into B2. The proficiency groups should therefore be treated as approximate indicators rather than precise measures of CEFR levels. The single trilingual questionnaire format rules out testing invariance across languages. All items are worded positively and the ceiling effect is clear. The common-method variance problem noted in [13] therefore applies to this analysis in full. Invariance shows that the instrument works in the same way across groups, but it does not prove that what is measured is a purely substantive construct. The boundary estimate of the EFF disturbance suggests that the second-order model might also be described by a structure with fewer than four domains. The search for partial invariance relied on a data-driven decision based on the score test, so the non-invariance of item AI3 should be re-examined in an independent sample.

Conclusion

The study provides psychometric evidence not reported before for the 13-item scale of acceptance of AI-gamified adaptive learning, its measurement invariance across groups. On the basis of the second-order factor model, the scale showed full scalar invariance across English proficiency. Loadings and intercepts alike are equivalent in the A2, B1 and B2+ groups, and the latent mean of general acceptance does not differ between levels. Across year of study, metric invariance was supported and scalar invariance proved partial, with one item concerning AI feedback released. Within the partial model, the latent mean of general acceptance is modestly higher for upper-year students than for first-year students. The main invariance conclusions remained stable under the ΔCFI criterion in the four-factor model and in the ordinal WLSMV sensitivity analysis.

The practical implications are as follows. There is a full psychometric basis for comparing scale scores across English proficiency. Comparison across year of study is best carried out through latent means within the reported partial invariance model, and item AI3 should be treated with caution when compared across years on its own. In methodological terms the study demonstrates, on real data, the testing sequence for the invariance of a second-order model and the procedure of partial invariance.

Future studies should re-examine the scale in samples from several universities with a more balanced gender composition. Developing separate Kazakh, Russian and English versions would open the way to testing invariance across languages. Adding negatively worded items and an external criterion, for instance usage logs and learning outcomes, would allow a formal test of common-method variance. These steps would move the scale from structural evidence towards full validation.

Funding information

The study was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant AP26101671, Development of a methodology for building the professional competence of future foreign language teachers in the context of digital adaptive learning.

References

1. Davis F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology / F. D. Davis // *MIS Quarterly*. - 1989. - Vol. 13, No. 3. - P. 319-340. <https://doi.org/10.2307/249008>
2. Venkatesh V. User acceptance of information technology: Toward a unified view / V. Venkatesh, M. G. Morris, G. B. Davis, F. D. Davis // *MIS Quarterly*. - 2003. - Vol. 27, No. 3. - P. 425-478. <https://doi.org/10.2307/30036540>
3. From game design elements to gamefulness: Defining "gamification" / S. Deterding, D. Dixon, R. Khaled, L. Nacke // *Proceedings of the 15th International Academic MindTrek Conference*. - ACM, 2011. - P. 9-15. <https://doi.org/10.1145/2181037.2181040>
4. Standards for educational and psychological testing / American Educational Research Association, American Psychological Association, National Council on Measurement in Education. - Washington, DC : AERA, 2014. - 224 p.
5. Vandenberg R. J. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research / R. J. Vandenberg, C. E. Lance // *Organizational Research Methods*. - 2000. - Vol. 3, No. 1. - P. 4-70. <https://doi.org/10.1177/109442810031002>
6. Meredith W. Measurement invariance, factor analysis and factorial invariance / W. Meredith // *Psychometrika*. - 1993. - Vol. 58, No. 4. - P. 525-543. <https://doi.org/10.1007/BF02294825>
7. Millsap R. E. Statistical approaches to measurement invariance / R. E. Millsap. - New York : Routledge, 2011. - 275 p.
8. Putnick D. L. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research / D. L. Putnick, M. H. Bornstein // *Developmental Review*. - 2016. - Vol. 41. - P. 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
9. Byrne B. M. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance / B. M. Byrne, R. J. Shavelson, B. Muthén // *Psychological Bulletin*. - 1989. - Vol. 105, No. 3. - P. 456-466. <https://doi.org/10.1037/0033-2909.105.3.456>
10. Chen F. F. Teacher's corner: Testing measurement invariance of second-order factor models / F. F. Chen, K. H. Sousa, S. G. West // *Structural Equation Modeling*. - 2005. - Vol. 12, No. 3. - P. 471-492. https://doi.org/10.1207/s15328007sem1203_7
11. Cheung G. W. Evaluating goodness-of-fit indexes for testing measurement invariance / G. W. Cheung, R. B. Rensvold // *Structural Equation Modeling*. - 2002. - Vol. 9, No. 2. - P. 233-255. https://doi.org/10.1207/S15328007SEM0902_5

12. Chen F. F. Sensitivity of goodness of fit indexes to lack of measurement invariance / F. F. Chen // Structural Equation Modeling. - 2007. - Vol. 14, No. 3. - P. 464-504. <https://doi.org/10.1080/10705510701301834>
13. Acceptance of AI-gamified adaptive learning in EFL: Bifactor evidence for essential unidimensionality and the limits of subscale scoring / A. Zhorabekova, M. Duisenova, M. Kurmasheva, A. Yesnazar // Frontiers in Education. - 2026. - Vol. 11. - Article 1929123. <https://doi.org/10.3389/educ.2026.1929123>
14. Hu L. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives / L. Hu, P. M. Bentler // Structural Equation Modeling. - 1999. - Vol. 6, No. 1. - P. 1-55. <https://doi.org/10.1080/10705519909540118>
15. Satorra A. A scaled difference chi-square test statistic for moment structure analysis / A. Satorra, P. M. Bentler // Psychometrika. - 2001. - Vol. 66, No. 4. - P. 507-514. <https://doi.org/10.1007/BF02296192>
16. Wu H. Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes / H. Wu, R. Estabrook // Psychometrika. - 2016. - Vol. 81, No. 4. - P. 1014-1045. <https://doi.org/10.1007/s11336-016-9506-0>
17. Rosseel Y. lavaan: An R package for structural equation modeling / Y. Rosseel // Journal of Statistical Software. - 2012. - Vol. 48, No. 2. - P. 1-36. <https://doi.org/10.18637/jss.v048.i02>
18. Jorgensen T. D. semTools: Useful tools for structural equation modeling (R package version 0.5-6) / T. D. Jorgensen, S. Pornprasertmanit, A. M. Schoemann, Y. Rosseel, 2022. <https://CRAN.R-project.org/package=semTools>
19. Kline R. B. Principles and practice of structural equation modeling / R. B. Kline. - 4th ed. - New York : Guilford Press, 2016. - 534 p.
20. Teo T. Unpacking teachers' acceptance of technology: Tests of measurement invariance and latent mean differences / T. Teo // Computers & Education. - 2014. - Vol. 75. - P. 127-135. <https://doi.org/10.1016/j.compedu.2014.01.014>

А.Н. Жорабекова, М.М. Дуйсенова, М.Б. Курмашева, А.Ж. Есназар

ЖИ-ГЕЙМИФИКАЦИЯЛАНҒАН АДАПТИВТІ ОҚЫТУДЫ ҚАБЫЛДАУ ШКАЛАСЫНЫҢ АҒЫЛШЫН ДЕҢГЕЙІ МЕН ОҚУ КУРСЫ БОЙЫНША ӨЛШЕМ ИНВАРИАНТТЫЛЫҒЫ

Жасанды интеллект пен геймификацияны біріктіретін адаптивті оқыту құралдарын студенттердің қабылдауы әдетте көпөлшемді сауалнамалармен өлшенеді. Мұндай сауалнаманың әртүрлі топтарда бірдей өлшемдік қасиет көрсететіні сирек тексеріледі. Зерттеудің мақсаты ЖИ-геймификацияланған адаптивті оқытуды қабылдаудың 13 пункттік шкаласының ағылшын тілін меңгеру деңгейі мен оқу курсы бойынша өлшем инварианттылығын тексеру және инварианттылық расталған жағдайда латентті орташа айырмашылықтарды бағалау болды. Қазақстандық педагогикалық университеттің 401 EFL студентінен бұрын жиналған деректер көптоптық растаушы факторлық талдау арқылы талданды. Негізгі модель ретінде екінші реттік факторлық модель алынды, сезімталдық талдауында төрт корреляцияланған факторлы модель MLR және WLSMV бағалаушыларымен пайдаланылды. Ағылшын деңгейі бойынша A2, B1 және B2+ топтарында шкала толық скалярлық инварианттылық көрсетті, CFI төмендеуі .003-тен аспады, χ^2 айырмасы бірде-бір кезеңде маңызды болмады, ал жалпы қабылдаудың латентті орташасы деңгейлер арасында ерекшеленбеді. Оқу курсы бойынша бірінші курс пен жоғары курстар арасында метрикалық инварианттылық расталды, скалярлық инварианттылық ішінара болып шықты. ЖИ кері байланысына қатысты бір пункттің интерсепті босатылғанда шектеулі және еркін модельдер маңызды түрде ерекшеленбеді. Ішінара инварианттылық моделі шеңберінде жоғары курс студенттерінің жалпы қабылдауының латентті орташасы бірінші курстан шамамен 0,3 стандартты ауытқуға жоғары. Нәтижелер ағылшын деңгейі бойынша салыстыруға толық, ал оқу курсы бойынша латентті орташаларды салыстыруға ішінара психометриялық негіз бар екенін көрсетеді. Зерттеудің шектеулері бір университеттің ыңғайлы іріктемесіне, деңгейдің өзін-өзі бағалау арқылы жиналуына, төбе эффектісіне және ортақ әдіске байланысты дисперсияның ықтималдығына қатысты.

Түйін сөздер: өлшем инварианттылығы, көптоптық растаушы факторлық талдау, екінші реттік факторлық модель, технологияны қабылдау, жасанды интеллект, геймификация, адаптивті оқыту, шетел тілін оқыту, психометрика.

А.Н. Жорабекова, М.М. Дуйсенова, М.Б. Курмашева, А.Ж. Есназар

ИЗМЕРИТЕЛЬНАЯ ИНВАРИАНТНОСТЬ ШКАЛЫ ПРИНЯТИЯ ИИ-ГЕЙМИФИЦИРОВАННОГО АДАПТИВНОГО ОБУЧЕНИЯ ПО УРОВНЮ АНГЛИЙСКОГО ЯЗЫКА И КУРСУ ОБУЧЕНИЯ

Принятие студентами адаптивных средств обучения, сочетающих искусственный интеллект и геймификацию, обычно измеряется многомерными опросниками, однако то, насколько одинаковые измерительные свойства такие опросники показывают в разных группах студентов, проверяется редко. Цель исследования состояла в проверке измерительной инвариантности 13-пунктовой шкалы принятия ИИ-геймифицированного адаптивного обучения по уровню владения английским языком и курсу обучения и в оценке различий латентных средних там, где инвариантность подтверждается. Ранее собранные данные 401 студента EFL казахстанского педагогического университета были проанализированы методом многогруппового конфирматорного факторного анализа. Основной служила факторная модель второго порядка, а модель четырёх коррелирующих факторов с оценками MLR и WLSMV использовалась для анализа чувствительности. По уровню английского языка в группах A2, B1 и B2+ шкала показала полную скалярную инвариантность, снижение CFI не превышало .003, разность χ^2 была незначима на всех этапах, а латентное среднее общего принятия между уровнями не различалось. По курсу обучения при сравнении первокурсников со студентами второго-четвёртого курсов подтверждена метрическая инвариантность, скалярная инвариантность оказалась частичной. После освобождения интерсепта одного пункта об обратной связи ИИ модели значимо не различались. В рамках частичной инвариантной модели латентное среднее общего принятия у студентов старших курсов выше, чем у первокурсников, примерно на 0,3 стандартного отклонения. Результаты дают полное психометрическое основание для сравнений по уровню языка и частичное основание для сравнения латентных средних по курсу обучения. Ограничения связаны с удобной выборкой одного университета, самооценкой уровня языка, эффектом потолка и возможностью дисперсии, обусловленной общим методом сбора данных.

Ключевые слова: измерительная инвариантность, многогрупповой конфирматорный факторный анализ, факторная модель второго порядка, принятие технологий, искусственный интеллект, геймификация, адаптивное обучение, обучение иностранному языку, психометрика.

References

1. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
2. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
3. Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference* (pp. 9–15). ACM. <https://doi.org/10.1145/2181037.2181040>
4. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.
5. Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/109442810031002>
6. Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
7. Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

8. Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
9. Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. <https://doi.org/10.1037/0033-2909.105.3.456>
10. Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, 12(3), 471–492. https://doi.org/10.1207/s15328007sem1203_7
11. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
12. Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
13. Zhorabekova, A., Duisenova, M., Kurmasheva, M., & Yesnazar, A. (2026). Acceptance of AI-gamified adaptive learning in EFL: Bifactor evidence for essential unidimensionality and the limits of subscale scoring. *Frontiers in Education*, 11, Article 1929123. <https://doi.org/10.3389/educ.2026.1929123>
14. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
15. Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. <https://doi.org/10.1007/BF02296192>
16. Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>
17. Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
18. Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling (R package version 0.5-6). <https://CRAN.R-project.org/package=semTools>
19. Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
20. Teo, T. (2014). Unpacking teachers' acceptance of technology: Tests of measurement invariance and latent mean differences. *Computers & Education*, 75, 127–135. <https://doi.org/10.1016/j.compedu.2014.01.014>

Information about the authors:

Zhorabekova Ainur Nazimbekovna - PhD, associate professor, M.Auezov South Kazakhstan University, Tauke khan avenue 5, 160012, Shymkent, Kazakhstan, e-mail: ainur.zhorabekova@aeuzov.edu.kz

Duisenova Marzhan Moldakassymovna (corresponding author) - corresponding author, PhD, acting professor O.Zhanibekov South Kazakhstan Pedagogical University, A. Baitursynov 13, 160009, Shymkent, Kazakhstan, e-mail: duisenova.marzhan@okmpu.kz

Kurmasheva Moldir Bizhigitovna - doctoral student of the educational program 8D01703 - “Teacher training in foreign language: two foreign languages”, O.Zhanibekov South Kazakhstan Pedagogical University, A. Baitursynov 13, 160009, Shymkent, Kazakhstan, e-mail: kurmasheva.moldir@okmpu.kz

Yesnazar Assel - PhD, senior lecturer, O.Zhanibekov South Kazakhstan Pedagogical University, A. Baitursynov 13, 160009, Shymkent, Kazakhstan, e-mail: asel.esnazar@mail.ru

Авторлар туралы мәліметтер:

Жорабекова Айнур Назимбековна - PhD, қауымдастырылған профессор, М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент қаласы, Тәуке хан даңғылы 5, 160012, Шымкент қ. Қазақстан, e-mail: ainur.zhorabekova@aeuzov.edu.kz

Дуйсенова Маржан Молдакасымовна (автор-корреспондент) - хат-хабар авторы, PhD, профессор м.а. Ө. Жәнібеков атындағы Оңтүстік Қазақстан педагогикалық университеті, А. Байтұрсынов 13, 160009, Шымкент қ. Қазақстан, e-mail: duisenova.marzhan@okmptu.kz

Курмашева Молдир Бижигитовна - 8D01703- Шетел тілі: екі шетел тілі педагогін даярлай білім беру бағдарламасының докторанты, Ө.Жәнібеков атындағы Оңтүстік Қазақстан педагогикалық университеті, А. Байтұрсынов 13, 160009, Шымкент қ., Қазақстан, e-mail: kurmasheva.moldir@okmptu.kz

Есназар Асель Жаңабергенқызы - PhD, аға оқытушы, Ө.Жәнібеков атындағы Оңтүстік Қазақстан педагогикалық университеті, Қазақстан, А. Байтұрсынов 13, 160009, Шымкент қ. Қазақстан, e-mail: asel.esnazar@mail.ru

Сведение об авторах:

Жорабекова Айнур Назимбековна, - PhD, ассоциированный профессор, Южно-Казахстанский университет им. М.Ауезова, город Шымкент, проспект Тауке хана 5, 160012, г. Шымкент, Казахстан, e-mail: ainur.zhorabekova@auuezov.edu.kz

Дуйсенова Маржан Молдакасымовна (автор-корреспондент) - автор для корреспонденции, PhD, и.о профессора Южно-Казахстанский педагогический университет им. Ө. Жәнібеков, А. Байтұрсынов 13, 160009, г. Шымкент, Казахстан, e-mail: duisenova.marzhan@okmptu.kz

Курмашева Молдир Бижигитовна - докторант образовательной программы 8D01703- Подготовка педагога по иностранному языку: два иностранных языка, Южно-Казахстанский педагогический университет имени Ө. Жәнібеков, А. Байтұрсынов 13, 160009, г. Шымкент, Казахстан, e-mail: kurmasheva.moldir@okmptu.kz

Есназар Асель Жаңабергенқызы - PhD, ст.преподаватель, Южно-Казахстанский педагогический университет имени Ө. Жәнібеков, А. Байтұрсынов 13, 160009, г. Шымкент, Казахстан, e-mail: asel.esnazar@mail.ru

M.E. Ospanaly^{1*}, A.K. Kalieva², N. Temur³

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan

²M. Auezov Institute of Literature and Art, Almaty, Kazakhstan

³Gazi University, Ankara, Turkey

*e-mail: ospanaly.kz@gmail.com

¹ORCID 0000-0002-2125-9287, ²ORCID 0000-0003-2645-0918,

³ORCID 0000-0002-8052-1927

EVALUATING AI RESPONSES IN LITERATURE TEACHING: THE LITERARY INTERIOR IN KAZAKH PROSE

This article examines the methodological potential of generative language models for developing philology students' ability to analyse and interpret the poetics of the interior in literary texts. The empirical basis of the study is an initial dialogue with Google Gemini held in February 2026 and the responses to the same prompts put to Gemini, ChatGPT and Claude eight months later. The models' responses were compared against predefined criteria, and their statements about the interior in Mukhtar Magauin's novel *Zharmak* were checked against the text of the novel and a published literary review. The analysis showed that all the models were useful as a source of methodological guidance. The analytical questions they proposed are consistent with the theories of literary space developed by Yuri Lotman, Gaston Bachelard and Alexander Chudakov, and they used the conceptual vocabulary of the traditional Kazakh dwelling correctly. In the close analysis of a specific literary text, however, their reliability proved limited. Of Gemini's 11 claims, only 2 were fully and 6 partly confirmed. ChatGPT confined itself to directing the user back to the text, while Claude drew on web sources. The novel's key spatial details, which none of the models quoted, emerged only through close reading. The study produced a three-stage model for teaching work with artificial intelligence in university philology courses and a seminar tasks that develops students' critical analysis skills.

Keywords: generative artificial intelligence, large language models, teaching literature, literary interior, literary space, Kazakh literature.

Introduction

Whether philology students will use generative language models is no longer an open question. They already do. For a literature teacher the more useful question is at which point in working with a literary text the model helps a student think and at which point it begins to do the thinking instead. University responses so far swing between prohibition and quiet acceptance, and neither gives teachers a method.

The research picture is mixed. A review of 48 peer-reviewed studies of large language models in literary studies identifies eight areas of use, from metaphor interpretation to translation, and four persistent limitations: bias in training data, dependence on human checking, weak handling of long and complex narratives, and shallow reasoning [1]. The models are not poor readers, though. Expert judges rated GPT-4's interpretations of unfamiliar poetic metaphors above those written by college students [2]. In an undergraduate literature course, texts produced by ChatGPT outsourced student texts on setting, theme and tone, even though students trusted the human-written texts more [3].

Fluent output is not evidence of learning. In a randomised study, ChatGPT support raised essay scores but produced no gain in knowledge or transfer, an effect the authors call metacognitive laziness [4]. In a field experiment, unrestricted access to GPT-4 improved scores on practice problems but lowered scores on a later exam taken without AI by 17%, while a version that offered hints instead of finished solutions did no such harm [5]. Sycophancy is a separate problem: systems trained on human preference ratings tend to agree with the user at the expense of accuracy [6]. That is a real risk for a student who arrives with a thesis already in mind.

In Kazakhstan the question also has a regulatory side. The Concept for the Development of Artificial Intelligence for 2024–2029 calls for wider use of AI in university programmes and for continuous training of teaching staff [7]. The AI-Sana programme aims to build AI competencies among bachelor's, master's and doctoral students [8], and the Law on Artificial Intelligence sets out the principles of transparency and the priority of human will [9]. Local empirical evidence is still thin. A quasi-experiment with literature students found higher academic results when ChatGPT and other tools were built into classes on ethnocultural codes [10]. Surveys show ambivalence: students find ChatGPT useful yet worry about losing their own writing skills [11], and they believe AI develops technical skills better than interpersonal ones [12]. Culture matters as well. The values expressed by these models lean towards English-speaking and Western European cultures [13], and their scores drop noticeably on tasks involving the Kazakh language and regional knowledge [14].

The literary interior is a good place to test these risks. For Yuri Lotman, space in a literary work is a model of the world built from oppositions and boundaries, and crossing a boundary is an event [15]. Gaston Bachelard read the house, with its corners, chests and wardrobes, as a map of inner life [16]. Alexander Chudakov showed that the way a writer selects and presents objects is a poetics in its own right, one that characterises people and their world [17]. Analysing an interior therefore asks three things of a student: general analytical categories, attention to specific detail, and knowledge of the cultural code. Our expectation was that a model would readily supply the first and would be unreliable on the other two.

Existing studies compare groups of students or rate the quality of AI-generated texts. Few look at the dialogue itself, and we found no study in which the same dialogue about Kazakh literature was run through several models and repeated over time. This gap shapes the present article.

The aim of the study is to compare the epistemic behaviour of generative models in a research dialogue about interiors in Kazakh literature and to derive recommendations for university literature courses. **The objectives** are: 1) to describe what the models' responses do in the original dialogue and in its replications; 2) to check the models' claims about a specific novel against secondary sources and against the text; 3) to compare the models with each other and one model with itself over time; 4) to propose indicators for evaluating AI responses and a way of using them in class. **The object of the study** is the use of generative AI in university literature teaching. **The subject** is the function and reliability of language model responses in a dialogue about the literary interior.

Novelty. The study compares three models on identical prompts, checks their claims in two stages up to the primary source, and repeats the test for one model after eight months. **Practical value.** It offers operational indicators that a teacher can use to evaluate AI responses and a seminar task built on real model output. **Hypothesis.** The models are comparable in suggesting directions, categories and structure but differ substantially in epistemic behaviour: in their readiness to challenge the user's framing, in their use of sources, and in the reliability of what they say about a specific Kazakh text.

Materials and methods

Material. The starting point is a dialogue between the first author and a custom assistant called "Literature Interior designer", built on Google Gemini with the Gems feature, in February 2026. It contains four prompts: on the qualities of a reviewer who studies interiors in literature; on choosing a topic for an article about interiors in Kazakh literature; on narrowing the topic "The 'Khrushchyovka' vs. The Steppe: Post-Nomadic Displacement" into a thesis and an outline; and on Mukhtar Magauin's novel Zharmak. The dialogue was exported to a text document, and a fragment of one response that had been pasted twice was excluded.

Replication. On 23 September 2026 the same four prompts were given, word for word and in the same order, to three systems, each in a new chat (Table 1). Access conditions were not identical, and we take this into account when interpreting the results. In the first ChatGPT session a typing error merged the second prompt with the first, so the session was restarted and only the second session is analysed.

Table 1 - Conditions of the four sessions

Session	Date	System and model	Access conditions
Gemini-02	February 2026	Google Gemini, custom Gem; model version not recorded	Author's account; the Gem's own system instructions were active
Gemini-09	September 2026	Google Gemini, 3.5 Flash-Lite	Not signed in; response 2 displayed links to travel websites
ChatGPT	September 2026	OpenAI ChatGPT, default model (the interface does not name the version)	Not signed in; web search used in response 3
Claude	September 2026	Anthropic Claude, Opus 5.5 (Medium setting)	Author's account, incognito mode; the account had personal preferences asking the model to test the user's claims critically; web search used in response 4
Note: compiled by the authors			

Design and indicators. We used a comparative case study with qualitative content analysis. The unit of analysis is one model response to one prompt, sixteen responses in total. The indicators are defined operationally (Table 2), so the counts can be reproduced from the transcripts.

Table 2 - Operational definitions of the indicators

Indicator	Definition
Response length	Number of words in the model's responses
Kazakh authors and works named	Specific writers or texts the model proposed as material, not counting the novel named by the user
Reference to the corpus	A response in which the model asks which texts the work will rest on, or insists that the thesis be tested on specific texts first
Challenge to the framing	A response in which the model questions the user's wording, term or thesis
Opening praise	The first sentence of the response praises the user or the user's choice
Claims about Zharmak	Information about the novel's content, motifs or publication history given by the model beyond what the user said
Sources	Links, quotations or signs of a completed search accompanying the claims
Note: compiled by the authors	

Checking the claims took two stages. First, the claims about Zharmak were compared with a published review [18]. Second, they were compared with an electronic edition of the novel itself, about 57,000 words long and dated by the author "18.I–9.II, 2007, Prague" [19]. We searched it for key words (päter, kitap, söre, stol, tör, bosaga, aina, zhusan, körshi, teledidar, qolzhazba and others) and read each hit in context. Each claim was then coded as confirmed, partly confirmed (the fact is in the text but the interpretation or a detail belongs to the model), not confirmed, or evaluative.

Why these methods. A case study suits an exploratory question. We were interested less in how often something happens in a population than in how a model's answer helps or hinders a student. Repeating the prompts verbatim makes the comparison reproducible, and checking claims in two stages separates a claim that cannot be confirmed from one that is false.

Results and discussion

Overview. Table 3 gives the quantitative indicators for the four sessions.

Table 3 - Indicators of the models' responses

Indicator	Gemini-02	Gemini-09	ChatGPT	Claude
Response length, words	1844	1472	3978	2526
Kazakh authors and works named	3 (1 not found)	1	5	1
Reference to the corpus, responses out of 4	0	0	3	3
Challenge to the framing, responses out of 4	0	0	2	4
Opening praise, responses out of 4	3	1	1	0
Claims about Zharmak	11	5	0	8
Claims with a source	0	0	n/a	8
Note: compiled by the authors from the session transcripts				

All four sessions were useful in the same respects. Each model proposed research directions, analytical categories and a structure for the paper. The first answer from every model is essentially a list of questions to ask of an interior, and nearly every question maps onto classical theory, although only Claude names that theory. “Who sits where, who stands by the window” is Lotman’s spatial hierarchy and boundary [15]. “The room as a reflection of the character’s inner state” is Bachelard’s topography of intimacy [16]. “Anachronisms in the furnishings” and “materials that give away social status” are Chudakov’s object as characterisation [17]. All the models also supplied cultural vocabulary (tör, shanyraq, tekemet, syrmaq, tus kiiz, quraq körpe and others) and used it broadly correctly. The differences were in emphasis. ChatGPT gave the strictest methodological rule: read the room as carefully as the character, but never let the room say more than the text can support. Claude warned that an editor trained only on Western interiors will turn personal taste into a norm. Gemini-09 drew on the vocabulary of design magazines (Eames, wabi-sabi) and ended with a quotation from an interior designer whose origin we did not verify.

Epistemic behaviour. The main differences appeared in the third and fourth turns, when the user brought a framing of his own and then a specific novel. Both Gemini sessions accepted the “Khrushchyovka versus the steppe” framing without objection and expanded it into an outline. ChatGPT suggested dropping the “vs.” because the opposition oversimplifies the relationship, and noted that the word “displacement” assumes the conclusion in advance; it supported this with web sources on Soviet mass housing. Claude said the framing “skips the middle”, meaning forced sedentarisation, the famine of the early 1930s, dugouts and collective-farm houses, and that it romanticises a steppe which, for the generation of collectivisation, was a place of starvation. Both models insisted that the thesis be tested on two or three specific texts. In none of its eight responses did Gemini ask which texts the user had read, and in February it opened three answers with praise (“What a brilliant niche”, “a poignant and intellectually rich choice”, “the perfect literary anchor for your thesis”). This fits the pattern of sycophancy described in the literature [6].

Claims about Zharmak. Table 4 shows the results of the two-stage check.

Table 4 - Status of the models' claims about Zharmak after checking against the text

Status	Gemini-02 (n = 11)	Gemini-09 (n = 5)	ChatGPT (n = 0)	Claude (n = 8)
Confirmed	2	3	n/a	5
Partly confirmed	6	2	n/a	2
Not confirmed	2	0	n/a	1
Evaluative	1	0	n/a	0
Note: compiled by the authors; checked against the review [18] and the text of the novel [19]				

At the first stage, using the review, only two of Gemini's eleven February claims could be confirmed: the year of publication and the motif of the double, Murat and his double Marat [18]. The rest looked invented: the hero's flat, a library as the "last trench" of identity, books squeezed into a cramped room, "stale air" against "the scent of wormwood". Reading the novel changed the picture. The narrator lives in a "jadau, jarly päter", a shabby, poor flat in a housing estate, with a kitchen of "nebäri segiz sharshy metr" (a mere eight square metres), a narrow corridor and one room that serves as study, bedroom and guest room at once. Books cover three walls, "half on the shelves, half on the floor" [19]. Six of the model's claims were therefore coded as partly confirmed: the fact is in the text, but the interpretation ("the last trench of Kazakh identity") or the detail ("a dusty room", "books on the Horde and the Khanate") comes from the model. Two claims are not supported. The word "jusan" (wormwood) does not occur in the novel, and instead of noisy neighbours and street traffic the text describes silence and darkness in which not even the headlights of cars on the wide street can be seen.

This result matters more than a simple error count. Gemini's guesses were plausible for the genre. A scholar's cramped flat full of books is a typical setting in post-Soviet fiction about the intelligentsia, and the model guessed it. But the guess was presented as knowledge, without a single quotation or page number, and it was mixed with invented sensory details set in quotation marks as though they were quotations. For a student this is the hardest kind of error to spot, because a partial match with the text lends credibility to the invented parts as well. It is also telling that the review could not separate the lucky guesses from the inventions. Only reading the novel could.

Eight months later Gemini followed the same approach: five claims, none with a source. Three are confirmed (publication in the mid-2000s, political critique, books and manuscripts in a cramped room) and two partly confirmed; the label "magical realism", for instance, is the model's, although the novel does contain fantastic episodes. The length and tone of the answers changed. The attitude to sources did not. ChatGPT made no claims at all about the content of the novel. It suggested first building a table of the novel's spatial vocabulary and warned that the strongest argument might turn out not to be "Khrushchyovka versus the steppe" at all. Claude searched the web and correctly described the novel's frame: the narrator presents the text as someone else's manuscript and lists the titles he rejected ("Qily taghdyrlar", "Kölenke", "Synar", "Astar", "Sülde") before choosing Zharmak, "a part of the whole, its second half" [19]. It was wrong, however, in saying that the novel moves between the nineteenth century and the present, and it could not establish the year of publication, which it stated openly.

What the text itself offers. None of the models quoted the novel even once, though quotation is what separates literary analysis from paraphrase. The interior of Zharmak supports analysis along the lines of all three theories. The writing desk stands "in the tör", the place of honour in a yurt, now moved into a room in a standard apartment block. When the double enters the flat, he "stepped freely over the threshold", went "straight to the tör" and sat down at the owner's desk [19]. Crossing the threshold and taking the tör is an event in Lotman's sense: the double takes the host's place [15]. The room is described as tight as a bladder and entirely taken over by books, much like Bachelard's shell [16]. A thick candle on the desk takes the narrator back to his childhood in the aul, when lessons were learned by a home-made oil lamp; the choice of that object works as the kind of characterisation Chudakov wrote about [17]. The contrast is the double's palace, with chandeliers shaped like a shanyraq and a dome built in the image of a yurt's crown: the nomadic code moves from the dwelling into the decor of power. Interestingly, Claude's hypothesis that the standard flat becomes an "unwilling host" to the spatial order of the yurt finds direct support in the text, although Claude did not know this and said plainly that the thesis "stands or falls on your texts".

Interpretation in the light of other studies. These findings agree with the limitations that the review of LLMs in literary studies attributes to shallow reasoning and data bias [1], and they are consistent with evidence that models write fluent interpretations [2, 3]. The fluency is there; grounding in the text is not. The mechanism of cognitive offloading [4, 5] is visible in the very shape of the February dialogue. Every answer ended by offering the next ready-made product, a

paragraph, a table or a set of quotations, rather than a question for the student. ChatGPT and Claude behaved more like a tutor with guardrails, the condition that protected learning in the experimental evidence [5]. Cultural bias [13, 14] was milder than we expected. The cultural vocabulary was mostly correct, and the errors lay not in the words but in knowledge of a particular Kazakh text. For academic integrity the practical lesson is that a log of the student's dialogue with the model tells a teacher more than an AI-text detector, and detectors are also known to misjudge writing by non-native speakers [20]. Assessment should look at the process as well as the product [21].

Limitations. We examined one dialogue scenario, with one run per model. Model output is stochastic, and a second run could produce different figures. Access conditions differed (Table 1): the February Gem ran with its own system instructions, ChatGPT and Gemini-09 were used without signing in, and Claude was used in an account whose personal preferences ask the model to test the user's claims critically. That setting probably increased Claude's readiness to challenge the framing, so its score on this indicator should be read as an upper estimate. The coding was done by one researcher, and inter-rater agreement was not measured. The electronic text of the novel was not collated with a printed edition. The results should therefore be treated as grounded hypotheses for a study with students.

Conclusion

The hypothesis was confirmed. The models are comparable in what might be called work before the text: they suggest directions, analytical categories, a structure for the paper and cultural vocabulary. Their epistemic behaviour differs considerably. In both sessions, eight months apart, Gemini accepted the user's framing, did not ask which texts the user had read, and gave details about the novel without sources. Some of those details turned out to be good genre guesses and some were invented, and only reading the primary source told the two apart. ChatGPT avoided claims about the novel's content and directed the user to the text. Claude drew on the sources it found, challenged the framing and made one factual error.

For the theory and practice of educational measurement, the study proposes a set of operational indicators of the epistemic quality of an AI response (Table 2): the share of claims with a source, the share of claims confirmed by the primary text, how often the model refers the user to a corpus, and how often it challenges the user's framing. Teachers can use these indicators when choosing a tool, and researchers when planning an experiment.

Our practical recommendations fall into three stages. Before the text, AI is acceptable as a source of directions, questions to ask of an interior and cultural terms, provided these are checked against a dictionary. Instead of the text, that is, when the model makes claims about the content of a work, its answer should not be accepted without a page reference and a quotation that the student has found. After the text, the model is useful as an opponent that looks for objections to the student's draft but does not write the analysis. The material of this study yields a ready seminar task: students receive Gemini's February answer about Zharmak, check each of its eleven claims against the novel, and then rewrite the model's "ready-made paragraph" so that every point rests on a quotation. The task trains close reading and a critical attitude to AI at the same time, in line with the goals of the national AI Concept and the AI-Sana programme [7, 8] and with the principle of human agency in the UNESCO guidance [22].

The next step is a study with groups of philology students in which work across the three stages is compared with traditional analysis and the quality of students' interpretations of interiors is measured with a rubric of tested inter-rater reliability.

References

1. Mozaffari N., Mehrangmarani B., Khorramrouz A. Applications, performance, and research gaps of large language models in literary studies: a scoping review // *Human Behavior and Emerging Technologies*. – 2026. – Vol. 2026. – Art. 8695447. – <https://doi.org/10.1155/hbe2/8695447>
2. Ichien N., Stamenković D., Holyoak K. J. Large language model displays emergent ability to interpret novel literary metaphors // *Metaphor and Symbol*. – 2024. – Vol. 39, № 4. – P. 296–309. – <https://doi.org/10.1080/10926488.2024.2380348>
3. Bellot A. R., Gutiérrez-Colón Plana M., Baran K. A. Redefining literature education: the role of ChatGPT in undergraduate courses // *International Journal of Artificial Intelligence in Education*. – 2025. – Vol. 35, № 5. – P. 3185–3201. – <https://doi.org/10.1007/s40593-025-00497-3>
4. Fan Y., Tang L., Le H., Shen K., Tan S., Zhao Y., Shen Y., Li X., Gašević D. Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance // *British Journal of Educational Technology*. – 2025. – Vol. 56, № 2. – P. 489–530. – <https://doi.org/10.1111/bjet.13544>
5. Bastani H., Bastani O., Sungu A., Ge H., Kabakcı Ö., Mariman R. Generative AI without guardrails can harm learning: evidence from high school mathematics // *Proceedings of the National Academy of Sciences of the USA*. – 2025. – Vol. 122, № 26. – Art. e2422633122. – <https://doi.org/10.1073/pnas.2422633122>
6. Sharma M., Tong M., Korbak T. [et al.] Towards understanding sycophancy in language models // *The Twelfth International Conference on Learning Representations (ICLR 2024)*. – 2024. – <https://arxiv.org/abs/2310.13548>.
7. Об утверждении Концепции развития искусственного интеллекта на 2024–2029 годы: постановление Правительства Республики Казахстан от 24 июля 2024 года № 592 // ИПС «Әділет». – <https://adilet.zan.kz/rus/docs/P2400000592>.
8. Об утверждении Паспорта проекта инновационной акселерационной программы AI-Sana: приказ Министра науки и высшего образования Республики Казахстан от 3 июня 2025 года № 290 // ИС «Параграф». – URL: https://prg.kz/document/?doc_id=31706872.
9. Об искусственном интеллекте: Закон Республики Казахстан от 17 ноября 2025 года № 230-VIII // ИПС «Әділет». – URL: <https://adilet.zan.kz/rus/docs/Z2500000230>.
10. Tazhibayeva S., Kanatkerey A., Mamayeva G., Beisenuly Z., Buribayeva M., Kemberbay R. The impact of artificial intelligence applications on linguistic and literary connections among Turkic peoples based on ethnocultural codes // *Journal of Ethnic and Cultural Studies*. – 2026. – Vol. 13, № 3. – P. 207–228. – <https://doi.org/10.66815/ejecs/3042>
11. Bekturova M., Tulepova S., Zhaitapova A. Predicting Kazakhstani TEFL students' continuance intention towards using ChatGPT in academic writing // *Education and Information Technologies*. – 2025. – Vol. 30, № 9. – P. 11889–11909. – <https://doi.org/10.1007/s10639-024-13306-6>
12. Potluri R. M., Jumasseitova A. K., Mukashev Y. Artificial intelligence in Kazakhstan higher education: computing students' perceptions, engagement, and skill development // *International Journal of Educational Research Open*. – 2026. – Vol. 10. – Art. 100573. – <https://doi.org/10.1016/j.ijedro.2026.100573>
13. Tao Y., Viberg O., Baker R. S., Kizilcec R. F. Cultural bias and cultural alignment of large language models // *PNAS Nexus*. – 2024. – Vol. 3, № 9. – Art. pgae346. – <https://doi.org/10.1093/pnasnexus/pgae346>
14. Togmanov M., Mukhituly N., Turmakhan D., Mansurov J., Goloburda M., Sakip A., Xie Z., Wang Y., Syzdykov B., Laiyk N., Aji A. F., Kochmar E., Nakov P., Koto F. KazMMLU: evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan // *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers)*. – Vienna: ACL, 2025. – P. 14403–14416. – <https://doi.org/10.18653/v1/2025.acl-long.701>
15. Лотман Ю. М. Структура художественного текста. – М.: Искусство, 1970. – 384 с.
16. Башляр Г. Избранное: Поэтика пространства / пер. с франц. – М.: РОССПЭН, 2004. – 376 с.
17. Чудаков А. П. Слово – вещь – мир. От Пушкина до Толстого: очерки поэтики русских классиков. – М.: Современный писатель, 1992. – 320 с.
18. Әбікенұлы С. Мұхтар-Иманның мұңы немесе «Жармақ» // *Заң газеті*. – 2010. – 5 қаңтар. – URL: <https://abai.kz/post/3006>.
19. Мағауин М. Жармақ: роман // *Zharar.com*. – <https://zharar.com/index.php?action=item&do=shorttexts&id=20339>.

20. Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. GPT detectors are biased against non-native English writers // Patterns. – 2023. – Vol. 4, № 7. – Art. 100779. – <https://doi.org/10.1016/j.patter.2023.100779>

21. MLA-CCCC Joint Task Force on Writing and AI. Working paper: overview of the issues, statement of principles, and recommendations. – Modern Language Association; Conference on College Composition and Communication, 2023. – URL: <https://aiandwriting.hcommons.org/working-paper-1>

22. Miao F., Holmes W. Guidance for generative AI in education and research. – Paris: UNESCO, 2023. – 44 p. – URL: <https://unesdoc.unesco.org/ark:/48223/pf0000386693>

М.Е. Оспаналы, А.Қ. Қалиева, Н. Темур

ӘДЕБИЕТТІ ОҚЫТУДАҒЫ ЖИ ЖАУАПТАРЫН БАҒАЛАУ: ҚАЗАҚ ПРОЗАСЫНДАҒЫ ӘДЕБИ ИНТЕРЬЕР

Мақалада генеративті тілдік модельдердің филолог-студенттердің көркем мәтіндегі интерьер поэтикасын саралау және түсіндіру құзыретін қалыптастырудағы әдістемелік әлеуеті зерделенеді. Зерттеудің эмпирикалық негізін 2026 жылдың ақпан айында Google Gemini жүйесімен жүргізілген алғашқы диалог пен араға сегіз ай салып Gemini, ChatGPT және Claude модельдеріне қайта қойылған сәйкес сауалдар нәтижесі құрады. Модельдер генерациялаған жауаптар алдын ала белгіленген нақты өлшемдер бойынша салғастырылып, олардың М. Мағауиннің «Жармақ» романындағы интерьерге қатысты пайымдаулары шығарманың түпнұсқа мәтінімен және әдеби-сыни рецензиямен салыстырыла тексерілді. Талдау барысы қарастырылған барлық модельдердің әдіснамалық бағдар беру деңгейінде пайдалы болғанын көрсетті. Жасанды интеллект ұсынған талдау сұрақтары Ю. М. Лотман, Г. Башляр және А. П. Чудаковтың көркем кеңістік туралы тұжырымдамаларымен үндестік тауып, қазақ дәстүрлі баспанасының ұғымдық аппараты дұрыс қолданылды. Алайда нақты көркем мәтінді терең талдауда модельдердің сенімділік деңгейінің шектеулі екені анықталды. Gemini ұсынған 11 тұжырымның тек 2 тұжырымы толық, 6 тұжырымы ішінара расталды. ChatGPT қолданушыны бастапқы мәтінге қайта бағыттаумен шектелді, ал Claude веб-дереккөздерге сүйенді. Романның мәнді кеңістік детальдары модельдер арқылы емес, тек мәтінді мұқият оқу тәсілімен ғана айқындалды. Зерттеу нәтижесінде жоғары оқу орындарының филологиялық курстарына арналған жасанды интеллектпен жұмыс істеудің үш кезеңді оқыту үлгісі мен студенттердің сыни талдау дағдысын шыңдайтын семинар тапсырмалары әзірленді.

Түйін сөздер: генеративті жасанды интеллект; үлкен тілдік модельдер; әдебиетті оқыту; әдеби интерьер; көркем кеңістік; қазақ әдебиеті.

М.Е. Оспаналы, А.К. Калиева, Н. Темур

ОЦЕНКА ОТВЕТОВ ИИ В ПРЕПОДАВАНИИ ЛИТЕРАТУРЫ: ЛИТЕРАТУРНЫЙ ИНТЕРЬЕР В КАЗАХСКОЙ ПРОЗЕ

В статье изучается методический потенциал генеративных языковых моделей в формировании у студентов-филологов умения анализировать и интерпретировать поэтику интерьера в художественном тексте. Эмпирическую основу исследования составили первоначальный диалог с системой Google Gemini, проведённый в феврале 2026 г., и ответы на те же запросы, повторно заданные спустя восемь месяцев моделям Gemini, ChatGPT и Claude. Ответы моделей сопоставлены по заранее установленным критериям, а их суждения об интерьере в романе М. Мағауина «Жармақ» проверены по тексту произведения и литературно-критической рецензии. Анализ показал, что все рассмотренные модели оказались полезны на уровне методологических ориентиров. Предложенные ИИ вопросы для анализа согласуются с концепциями художественного пространства Ю. М. Лотмана, Г. Башляра и А. П. Чудакова, а понятийный аппарат традиционного казахского жилища использован корректно. Вместе с тем при углубленном анализе конкретного художественного текста надежность моделей оказалась ограниченной. Из 11 суждений Gemini полностью подтвердились лишь 2, частично 6. ChatGPT ограничился тем, что возвращал пользователя к исходному тексту, а Claude

опирался на веб-источники, и из 8 его утверждений подтвердились 5. При этом ни одна модель не привела конкретных цитат из романа. Значимые пространственные детали романа были выявлены не с помощью моделей, а только путем внимательного чтения текста. По итогам исследования разработаны трехэтапная модель обучения работе с искусственным интеллектом для филологических курсов вузов и семинарские задания, развивающие у студентов навыки критического анализа.

Ключевые слова: генеративный искусственный интеллект; большие языковые модели; преподавание литературы; литературный интерьер; художественное пространство; казахская литература.

References

1. Mozaffari, N., Mehrangmarani, B., & Khorramrouz, A. (2026). Applications, performance, and research gaps of large language models in literary studies: A scoping review. *Human Behavior and Emerging Technologies*, 2026, Article 8695447. <https://doi.org/10.1155/hbe2/8695447>
2. Ichien, N., Stamenković, D., & Holyoak, K. J. (2024). Large language model displays emergent ability to interpret novel literary metaphors. *Metaphor and Symbol*, 39(4), 296-309. <https://doi.org/10.1080/10926488.2024.2380348>
3. Bellot, A. R., Gutiérrez-Colón Plana, M., & Baran, K. A. (2025). Redefining literature education: The role of ChatGPT in undergraduate courses. *International Journal of Artificial Intelligence in Education*, 35(5), 3185–3201. <https://doi.org/10.1007/s40593-025-00497-3>
4. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
5. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), Article e2422633122. <https://doi.org/10.1073/pnas.2422633122>
6. Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askill, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2024). Towards understanding sycophancy in language models. In *The Twelfth International Conference on Learning Representations (ICLR 2024)*. <https://arxiv.org/abs/2310.13548>
7. Pravitel'stvo Respubliki Kazakhstan. (2024). *Ob utverzhdanii Kontseptsii razvitiya iskusstvennogo intellekta na 2024–2029 gody* [On approval of the Concept for the development of artificial intelligence for 2024–2029] (Postanovlenie No. 592 ot 24 iyulya 2024 goda). <https://adilet.zan.kz/rus/docs/P2400000592>
8. Ministerstvo nauki i vysshego obrazovaniya Respubliki Kazakhstan. (2025). *Ob utverzhdanii Pasporta proekta innovatsionnoy akseleratsionnoy programmy AI-Sana* [On approval of the project passport of the AI-Sana innovative acceleration programme] (Prikaz No. 290 ot 3 iyunya 2025 goda). https://prg.kz/document/?doc_id=31706872
9. Respublika Kazakhstan. (2025). *Ob iskusstvennom intellekte* [On artificial intelligence] (Zakon No. 230-VIII ot 17 noyabrya 2025 goda). <https://adilet.zan.kz/rus/docs/Z2500000230>
10. Tazhibayeva, S., Kanatkerey, A., Mamayeva, G., Beisenuly, Z., Buribayeva, M., & Kemberbay, R. (2026). The impact of artificial intelligence applications on linguistic and literary connections among Turkic peoples based on ethnocultural codes. *Journal of Ethnic and Cultural Studies*, 13(3), 207–228. <https://doi.org/10.66815/ejecs/3042>
11. Bekturova, M., Tulepova, S., & Zhaitapova, A. (2025). Predicting Kazakhstani TEFL students' continuance intention towards using ChatGPT in academic writing. *Education and Information Technologies*, 30(9), 11889–11909. <https://doi.org/10.1007/s10639-024-13306-6>
12. Potluri, R. M., Jumasseitova, A. K., & Mukashev, Y. (2026). Artificial intelligence in Kazakhstan higher education: Computing students' perceptions, engagement, and skill development. *International Journal of Educational Research Open*, 10, Article 100573. <https://doi.org/10.1016/j.ijedro.2026.100573>
13. Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), Article pgae346. <https://doi.org/10.1093/pnasnexus/pgae346>
14. Togmanov, M., Mukhituly, N., Turmakhan, D., Mansurov, J., Goloburda, M., Sakip, A., Xie, Z., Wang, Y., Syzdykov, B., Laiyk, N., Aji, A. F., Kochmar, E., Nakov, P., & Koto, F. (2025). KazMMLU:

Evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 14403–14416). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.acl-long.701>

15. Lotman, Yu. M. (1970). *Struktura khudozhestvennogo teksta* [The structure of the artistic text]. Iskusstvo.

16. Bashlyar, G. (2004). *Izbrannoe: Poetika prostranstva* [Selected works: The poetics of space]. ROSSPEN. (Original work published 1957)

17. Chudakov, A. P. (1992). *Slovo – veshch' – mir. Ot Pushkina do Tolstogo: ocherki poetiki russkikh klassikov* [Word – thing – world. From Pushkin to Tolstoy: Essays on the poetics of Russian classics]. Sovremennyy pisatel'.

18. Abikenuly, S. (2010, January 5). Mukhtar-Imannyn muny nemese «Zharmak» [Mukhtar-Iman's sorrow, or «Zharmak»]. *Zan gazeti*. <https://abai.kz/post/3006>

19. Magauin, M. (2007). *Zharmak: roman* [Zharmak: A novel]. Zharar.com. <https://zharar.com/index.php?action=item&do=shorttexts&id=20339>

20. Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), Article 100779. <https://doi.org/10.1016/j.patter.2023.100779>

21. MLA-CCCC Joint Task Force on Writing and AI. (2023). *MLA-CCCC Joint Task Force on Writing and AI working paper: Overview of the issues, statement of principles, and recommendations*. Modern Language Association; Conference on College Composition and Communication. <https://aiandwriting.hcommons.org/working-paper-1/>

22. Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>

Information about authors:

Ospanaly Madina Eldoskyzy (corresponding author) - doctoral student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: ospanaly.kz@gmail.com

Kalieva Almira - Candidate of Philological Sciences, M. Auezov Institute of Literature and Art, Almaty, Kazakhstan, e-mail: almira_kalieva.8@mail.ru

Temur Nezir - Doctor, Professor, Gazi University, Ankara, Turkey, e-mail: ntemur@gazi.edu.tr

Авторлар туралы мәлімет:

Оспаналы Мадина Елдоскызы (автор-корреспондент) - докторант, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, e-mail: ospanaly.kz@gmail.com

Қалиева Альмира Қайыртайқызы - филология ғылымдарының кандидаты, М. Әуезов атындағы Әдебиет және өнер институты, Алматы, Қазақстан, e-mail: almira_kalieva.8@mail.ru

Темур Незир - доктор, профессор, Гази университеті, Анкара, Түркия, e-mail: ntemur@gazi.edu.tr

Сведения об авторах:

Оспаналы Мадина Елдоскызы (автор-корреспондент) - докторант, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан, e-mail: ospanaly.kz@gmail.com

Калиева Альмира Кайыртаевна - кандидат филологических наук, Институт литературы и искусства имени М. Ауэзова, Алматы, Казахстан, e-mail: almira_kalieva.8@mail.ru

Темур Незир - доктор, профессор, Университет Гази, Анкара, Турция, e-mail: ntemur@gazi.edu.tr

A.U. Serikbayeva^{1*}, D.M. Akizhanova²¹Al-Farabi Kazakh National University, Almaty, Kazakhstan²L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

*e-mail: assiyaserikbayeva237@gmail.com

¹ORCID 0009-0002-7516-482X, ²ORCID 0000-0002-1300-0339

CAN GENERATIVE AI REPLACE HUMAN ESSAY RATERS? A PSYCHOMETRIC COMPARISON OF GPT-5.2, GEMINI, AND TEACHER RATINGS

Evaluating student writing is a crucial yet time-consuming process. Large Language Models (LLMs) and other recent advancements in generative AI provide potential solutions to support teachers in essay assessment. This study psychometrically evaluates large language models such as GPT-5.2 and Gemini with human raters to examine reliability of these solutions. All 55 essays were collected and assessed by GPT 5.2, Gemini and an experienced teacher using a four-criterion scoring rubric. The aim of this study is to psychometrically compare the results of human raters, GPT-5.2, and Gemini assessments on English essays, to determine their rating reliability, intercorrelation, rating differences, and the construct validity of the rating rubric. The results demonstrated that GPT-5.2 assigned the highest vocabulary scores, while Gemini produced the highest grammar scores. Overall, GPT-5.2 showed high agreement with teacher ratings and can be used as a reliable adjunct to essay grading. However, teacher professional judgment remains important when assessing complex content and contextual features. These findings support the use of generative AI as a supplementary tool for improving the efficiency and consistency of essay assessment in EFL contexts.

Keywords: automated essay scoring, Large Language Models (LLMs), GPT-5.2, Gemini, human rater, assessment; psychometric assessment.

Introduction

Writing assessment is a crucial component of language education. It is not only a tool for determining the level of linguistic competence, but also an important pedagogical process that allows to improve the quality of teaching, monitor the learning outcomes of students, and provide effective feedback [1].

Although traditional essay assessment is one of the primary methods for assessing students' writing competence, a number of issues remain regarding its reliability and fairness. First, human ratings are influenced by factors such as the assessor's experience, linguistic and cultural characteristics, and personal perceptions of the quality of written work. Such differences can lead to inconsistencies between assessors and variability in assessment outcomes [2].

Additionally, manually grading essays is time-and labour-intensive. Human evaluation rigor varies when grading essays. Such differences threaten the validity and fairness of writing assessment [3].

Advancements in Artificial Intelligence (AI) and large language models (LLMs) such as GPT-5.2 and Gemini offer a new opportunity for educational assessment.[2] These models are capable of analysing and generating natural language with high proficiency, also they can evaluate time consuming assignments and provide feedback [4].

Against this backdrop, large language models (LLMs) have been widely used in the field of automatic essay grading. Previous studies have shown that large language models such as GPT-5.2 can closely match the results of human graders, it has not yet been fully proven that they perform equally well across all grading criteria. Therefore, a comprehensive psychometric assessment of the essay assessment capabilities of large language models is one of the most pressing issues in contemporary educational research. While many studies have shown that GPT-5.2 or other large language models have a high level of agreement with human evaluation, several important gaps remain. First, most studies have considered only one large language model and have not compared

the evaluation results of multiple AI systems [4, 5, 6]. Moreover, there are very few empirical studies in the EFL (English as a Foreign Language) context in Kazakhstan comparing GPT-5.2, Gemini, and teacher assessments from a psychometric perspective [7, 8]. Therefore, the aim of this study is to psychometrically compare the results of human raters, GPT-5.2, and Gemini assessments on English essays, to determine their rating reliability, intercorrelation, rating differences, and the construct validity of the rating rubric. In this regard, the objectives of the study are

To compare the descriptive indicators of the Teacher, GPT-5.2 and Gemini ratings.

To determine the reliability and construct validity of the rating rubric.

To assess the inter-rater agreement and differences from a psychometric perspective.

Overall, these research questions aim to comprehensively evaluate the psychometric characteristics of GPT-5.2, Gemini, and teacher raters, and provide scientific evidence for the reliable and valid use of generative artificial intelligence in grading English essays.

Literature review

In recent years, the potential of large language models (LLMs) for assessing written work has been widely investigated. García-Varela et al. [9] studied the stability and fairness of the GPT-5.2 system in assessing essays and showed that the results of the AI assessment directly depend on the quality of the rubric used and the prompt structure. The results of the study showed that when assessing with simple instructions, the grades were less stable, but when using a specific rubric, assessment examples and a standardized assessment format, the assessment results were more stable and consistent. The authors proposed methods such as decision tables, rubric normalization and prompt engineering to improve the reliability of the assessment.

Moreover, Seßler et al. [10] compared the results of GPT-3.5, GPT-4, o1, LLaMA, and Mixtral models in assessing student essays in German with the ratings of 37 teachers. The study found that while the closed models (especially o1 and GPT-4) showed high agreement with the teacher ratings, the open models showed lower agreement. Furthermore, while the results of the models on linguistic criteria were closer to the teachers, differences remained on the content criteria.

Karaçeper and Kıray [11] investigated the ability of the ChatGPT-4o model to automatically grade essays in English and found that it only showed fair agreement with human raters. The authors concluded that although ChatGPT-4o was able to perform the grading quickly and objectively, it did not fully account for complex linguistic features and pedagogical considerations, and therefore could not fully replace human raters in high-stakes grading.

Wu et al. [12] found that an automated essay grading system based on GPT demonstrated high scoring consistency compared to human raters. Using the Many-Facet Rasch Measurement (MFRM) model, the authors showed that GPT grades, while having high internal consistency, differed from human graders on more complex criteria such as content. They also suggested that GPT could be used as a supplementary tool to enhance grading reliability, rather than as a complete replacement for human grading.

Within the Kazakhstani context, research on AI-based development and evaluation of Unified National Testing (UNT) items measurement was conducted by Abdrasilov et. al [8] that the most AI-generated tasks were considered suitable after expert review. To promote a human-in-the-loop approach to educational evaluation, the authors stressed that expert validation was still crucial.

Yermaganbetova et al. [13] studied the effectiveness of an AI-integrated educational platform in higher education and showed that artificial intelligence improved students' academic achievement, learning motivation, and self-regulated learning skills through automated assessment, personalized feedback, and adaptive learning. In the study, GPT was used primarily for question generation and formative feedback, while Gemini was used for rubric-based assessment of practical work. The authors also noted that AI assessment should be viewed as a tool to support, rather than completely replace, the teacher's professional judgment.

A review of the literature shows that most previous studies have compared a single large language model, such as GPT, with human raters. However, there is still a lack of research that compares multiple modern generative AI models, such as GPT-5.2 and Gemini, with teacher ratings on the basis of psychometric performance simultaneously. Furthermore, inter-rater agreement,

internal reliability, and criterion-referenced differences in ratings have not been comprehensively addressed.

Materials and research methods

This research employed a quantitative research design to compare human ratings with Large Language Models (LLMs) such as Chat GPT 5.2 and Gemini. The study involved 55 ninth-grade students from a comprehensive school in Almaty. Students' ages ranged from 15 to 16 years. The study participants were assessed on their writing and presentation of an essay on a given topic. The study used a convenience sampling method. Students completed **an essay writing task** entitled *"The Role of Education in Shaping a Successful Future"*. Students completed a minimum of 150 and a maximum of 180 words essay on the assigned topic during a regular English writing lesson. The written essay was then graded by the teacher using Chat GPT 5.2 and Gemini. The teacher was unaware of Large Language Models' grading during the grading process. Participants were assessed according to four criteria: organization, vocabulary, grammar, task achievement. Each criterion was scored on a five-point scale, resulting in a maximum total score of 20 points. All essays were evaluated by three independent raters: an experienced English teacher, GPT 5.2 and Gemini. All anonymous essays were first gathered.

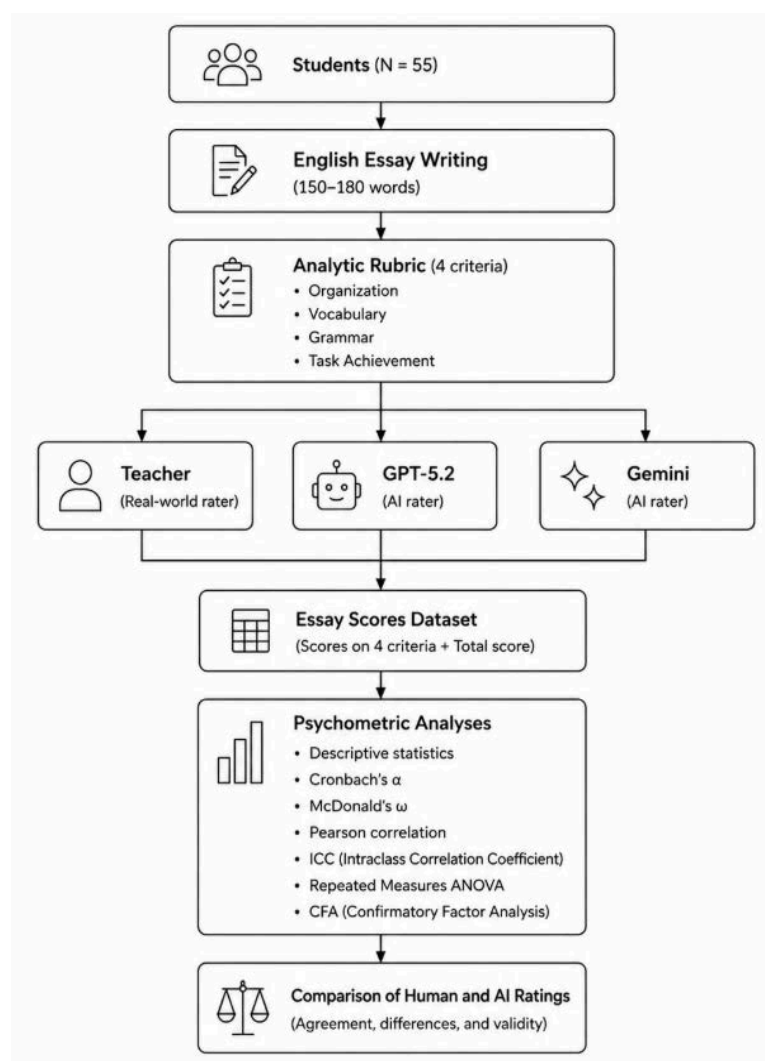


Figure 1 - Workflow of essay evaluation and psychometric assessment procedure

To ensure consistency across AI-based evaluations, the same prompt was used for GPT-5.2 and Gemini. The complete prompt is provided in Appendices A to D. The scores were then imported into Excel and statistical analysis was calculated. Statistical analyses were conducted using Jamovi (version 2.6.45.0). In order to summarize all essay scores descriptive statistics: mean and standard deviation also Cronbach's alpha and McDonald's omega were calculated. Pearson correlation coefficients were calculated to determine relationships among three raters. To examine the construct validity of the writing assessment rubric, Confirmatory Factor Analysis (CFA) was conducted.

Ethical Considerations

Informed consent was obtained from all participants before the study was conducted. Since the participants were school students, permission was also obtained from the school administration by informing their parents. Participants were informed about the purpose of the study and were given full information about their right to withdraw from participation at any time. Participants were informed that their personal information would be kept anonymous and confidential throughout the study.

Results and discussion

Table 1 - Descriptive Statistics of Essay Scores by Rater

Criterion	Teacher M (SD)	GPT-5.2 M (SD)	Gemini M (SD)
Organisation	3.96 (1.03)	3.96 (0.97)	3.78 (1.31)
Vocabulary	3.78 (0.96)	4.17 (0.75)	3.59 (1.10)
Grammar	3.82 (0.96)	3.66 (0.69)	4.14 (0.64)
Content	3.95 (1.02)	3.92 (1.22)	3.94 (0.66)
Total	15.50 (3.71)	15.60 (3.11)	15.40 (3.04)

Table 1 demonstrates the descriptive statistics of the three raters (teacher, GPT-5.2 and Gemini). According to the results, the highest mean score for the vocabulary criterion was shown by GPT-5.2 ($M = 4.17$, $SD = 0.75$). The highest mean score for the grammar section was observed in Gemini ($M = 4.14$, $SD = 0.64$). The teacher's ratings were relatively similar across the four assessment criteria. As for the overall composite scores, the average results of the three raters were very close to each other (teacher: $M = 15.50$; GPT-5.2: $M = 15.60$; Gemini: $M = 15.40$). These results indicate that the mean scores of the raters did not differ significantly and that their assessment processes were generally consistent.

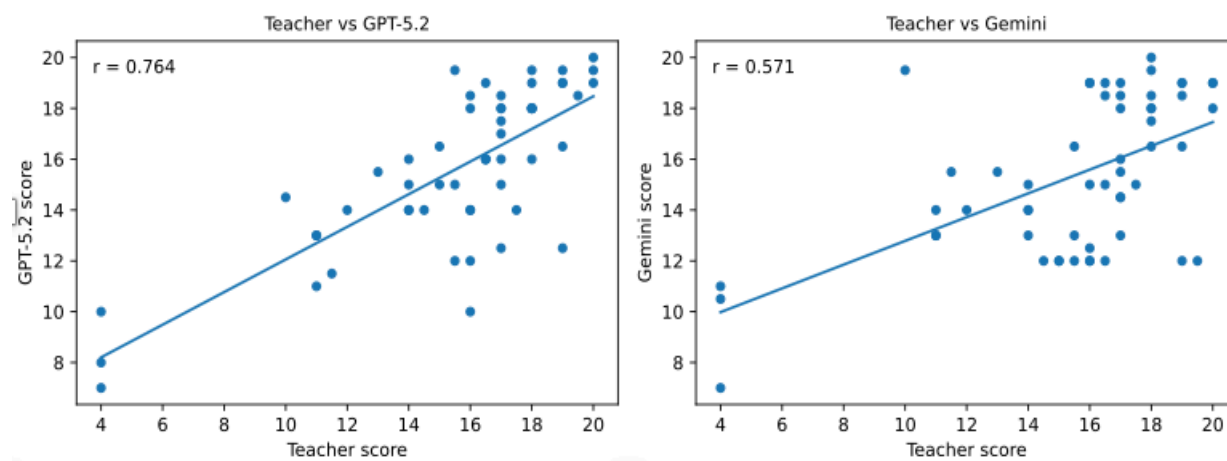
Table 2 - Internal Consistency of the Essay Rating Rubric

Rater	Cronbach's α	McDonald's ω
Teacher	0.947	0.947
GPT-5.2	0.846	0.870
Gemini	0.825	0.857

The highest internal consistency was found in human ratings $\alpha = 0.947$. This suggests that the rubric performed very consistently in human scoring. In addition, GPT-5.2 also used the four criteria consistently, with $\alpha = 0.846$. However, its internal consistency was lower than that of the teacher. Additionally, Gemini demonstrated good reliability (Cronbach's $\alpha = 0.825$). However, its internal consistency is slightly lower compared to GPT and human rater.

Table 3 - Pearson Correlation Matrix Among Essay Ratings

Variable	1	2	3
1. Teacher	-		
2. GPT-5.2	0.764***	-	
3. Gemini	0.571***	0.570***	-

**Figure 2 - Correlation between teacher ratings and AI-generated essay scores (GPT-5.2 and Gemini)**

Pearson correlation analysis was conducted to determine the relationship between human score, GPT-5.2, and Gemini scores. The results revealed a strong positive correlation between teacher and GPT-5.2 scores ($r = .764$, $p < .001$). This indicates that GPT-5.2 scores are highly correlated with teacher scores. Meanwhile, there was a moderate positive correlation between teacher and Gemini ($r = .571$, $p < .001$), as well as GPT-5.2 and Gemini ($r = .570$, $p < .001$). This indicates that Gemini scores are somewhat consistent with teacher scores, but are less correlated than GPT-5.2.

Table 4 - Repeated Measures ANOVA Comparing Essay Scores Across Raters

Source	SS	df	MS	F	<i>p</i>	Partial η^2
Rater	1.61	2	0.806	0.201	.818	.004
Error	432.72	108	4.007	—	—	—

Note. SS = Sum of Squares; MS = Mean Square; Partial η^2 = partial eta squared

A Repeated Measures ANOVA was conducted to determine whether there was a statistically significant difference between the total essay scores of the teacher, GPT-5.2, and Gemini raters. The results of the analysis showed that there was no statistically significant difference between the mean scores of the raters ($F(2,108) = 0.201$, $p = .818$, partial $\eta^2 = .004$). This suggests that the total essay scores of the teacher, GPT-5.2, and Gemini raters were very similar and that the type of rater did not significantly affect the results. Overall, the results obtained indicate that the teacher, GPT-5.2, and Gemini raters rated the essays at a similar level.

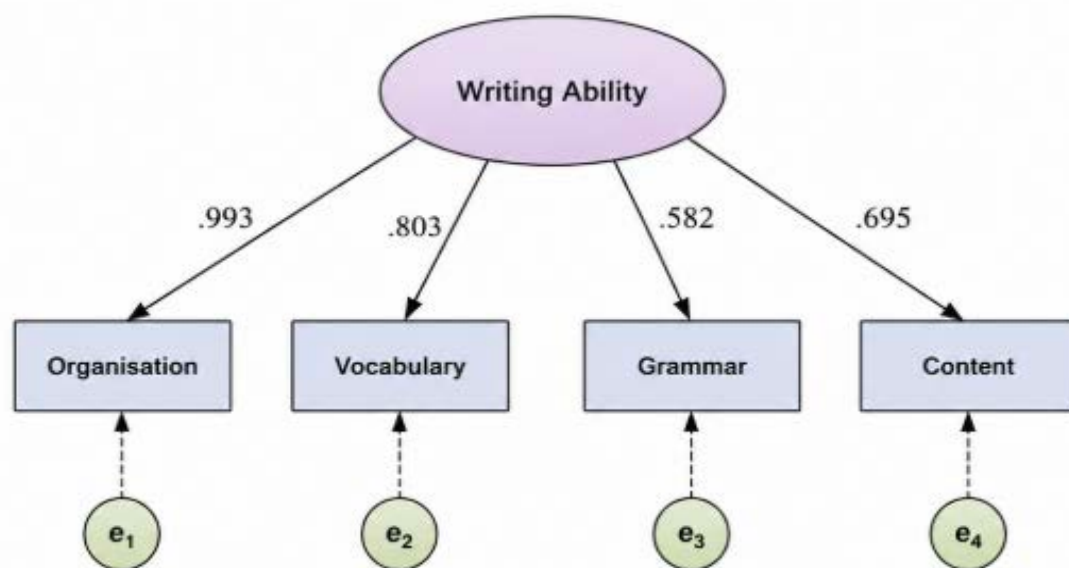
Table 5 - Intraclass Correlation Coefficients (ICC) Between Human and AI Raters

Comparison	ICC	95% CI	F	p
Teacher vs GPT-5.2	0.756	0.615-0.850	7.09	< .001
Teacher vs Gemini	0.563	0.351-0.720	3.54	< .001

In order to investigate inter-rater agreement Interclass Correlation Coefficients (ICC) were calculated for the teacher GPT 5.2 and teacher Gemini comparisons. The results indicate that good agreement between teacher and GPT 5.2 ($ICC = .756$, 95% CI [.615, .850], $p < .001$). However, teacher and Gemini showed moderate agreement ($ICC = .563$, 95% CI [.351, .720], $p < .001$). These findings suggest that GPT 5.2 can be more reliable tool for essay scoring.

Table 6 - Confirmatory Factor Analysis Results for the Writing Ability Construct

Indicator	Standardized Loading (λ)	SE	Z	p
Organisation	0.993	0.071	14.06	< .001
Vocabulary	0.803	0.064	12.52	< .001
Grammar	0.582	0.057	10.24	< .001
Content	0.695	0.070	9.88	< .001

**Figure 3 - Confirmatory Factor Analysis Model**

Confirmatory Factor Analysis (CFA) results showed that the four criteria of the writing assessment rubric (Organization, Vocabulary, Grammar, and Content) all measure a common latent construct called “Writing Ability”. All factor loadings were statistically significant ($p < .001$), which proves that these criteria make an important contribution to the assessment of writing ability. Comparing the factor loadings, the Organization indicator had the highest loading ($\lambda = 0.993$), which means that it is the strongest indicator in describing writing ability. It was followed by Vocabulary ($\lambda = 0.803$) and Content ($\lambda = 0.695$). The Grammar indicator showed the lowest factor loading ($\lambda = 0.582$), although its contribution was also statistically significant. In terms of model fit indices, the values of $CFI = 0.935$ and $SRMR = 0.043$ indicated an acceptable fit of the model.

However, TLI = 0.806 and RMSEA = 0.262 this result can be explained by the small sample size in the study ($N = 55$) and the fact that the model consists of only four indicators.

Overall, the study showed that there was no statistically significant difference between the mean scores of the three raters ($p = .818$). This suggests that GPT-5.2 and Gemini systems can produce results close to the teacher's overall grading level. The present findings are consistent with those reported by Seßler et al. [10] and Wu et al. [12] that Large Language Models (LLMs) showed the highest agreement with human raters and good internal consistency. This may be explained by current large language models (LLMs) can produce results that are close to those of teachers when using an analytical assessment rubric.

Furthermore, Chat GPT demonstrated the highest vocabulary scores among all raters. According to results GPT is good at recognizing lexical diversity and word selection, similar to the results reported by Kasneci et al. [4], and García-Varela et al. [9]. This may be attributed that rubric-guided ChatGPT assessment can provide stable and consistent evaluations across analytic writing criteria.

Moreover, Gemini demonstrated higher scores in grammar. A plausible explanation is that large language models are able to effectively learn grammatical structures and language patterns, they can demonstrate high accuracy in detecting grammatical errors. This finding aligns with Mizumoto & Eguchi [5] who reported that Chat GPT has superior ability to detect grammatical errors and assess linguistic accuracy in assessing written work.

Internal consistency results also support this conclusion. Teacher ratings showed the highest reliability ($\alpha = .947$), while GPT-5.2 ($\alpha = .846$) and Gemini ($\alpha = .825$) both had reliability coefficients above .80. This indicates that both AI systems were able to consistently use the analytical rubric. These results are consistent with the findings of a literature review by Lo [6] and a paper published by Kasneci and colleagues [4] These studies show that while modern large language models are able to consistently apply assessment criteria, the professional judgment and assessment stability of experienced teachers still remain at a high level.

The results align with previous research by Abdrasilov et. al [8] who investigated the use of AI assisted Large Language Models for generating tasks for the Unified National Testing (UNT). The research work focused on automatic generation of tasks and their psychometric quality rather than essay scoring. Their findings support human-in-loop approach which is consistent with present study that GPT provides teacher-like results and AI speeds up grading, but human judgment is important because AI does not always fully understand the context. This can be explained by Large Language Models (LLMs) cannot always correctly understand the creative nature of writing, cultural context, or unique responses.

Limitations

First, the limited number of participants ($N = 55$) may have affected the CFA model. High RMSEA in this study indicated that the single-factor model did not fully fit. This may be due primarily to the small sample size and the use of only four indicators. Second, the essays were collected from only one high school and only one task was included in the sample. Third, two large language models, GPT 5.2 and Gemini, were used. Therefore, future studies should increase the sample size and re-examine the structural validity of the model with broader comparison of AI-assisted essay assessment.

Conclusion

Overall, the results of the study showed that GPT-5.2 and Gemini systems can perform assessments that are close to human raters in grading English essays. In particular, GPT-5.2 showed a high level of agreement with teacher ratings. However, it was shown that artificial intelligence systems cannot fully replace the professional judgment of teachers in complex aspects such as understanding students' thinking, assessing creative features, and fully considering the contextual meaning of the text.

In conclusion it is appropriate to consider GPT-5.2 and other large language models not as a substitute for teacher judgment, but as an additional support tool that increases the efficiency and consistency of the assessment process.

Funding information

This research received no external funding.

References

1. Ma M. Exploring learning-oriented assessment in EAP writing classrooms: Teacher and student perspectives // *Language Testing in Asia*. - 2023. - T. 13. - Article 33. - <https://doi.org/10.1186/s40468-023-00249-x>
2. Han T., Zheng Y. AI and human scoring for postgraduate writing: Evaluating score reliability, variability, and rater behaviours // *Studies in Educational Evaluation*. - 2026. - T. 88. - Article 101572. - <https://doi.org/10.1016/j.stueduc.2026.101572>
3. Khamboonruang A. Detecting differential rater severity in a high-stakes EFL classroom writing assessment: A many-facets Rasch measurement approach // *PASAA*. - 2023. - T. 66. - C. 5–36.
4. Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Gasser U., Groh G., Günnemann S., Hüllermeier E., Krusche S., Kutyniok G., Michaeli T., Nerdel C., Pfeffer J., Poquet O., Sailer M., Schmidt A., Seidel T., Kasneci G. ChatGPT for good? On opportunities and challenges of large language models for education // *Learning and Individual Differences*. - 2023. - T. 103. - Article 102274. - <https://doi.org/10.1016/j.lindif.2023.102274>
5. Mizumoto A., Eguchi M. Exploring the potential of using an AI language model for automated essay scoring // *Research Methods in Applied Linguistics*. - 2023. - T. 2, № 2. - Article 100050. - <https://doi.org/10.1016/j.rmal.2023.100050>
6. Lo C. K. What is the impact of ChatGPT on education? A rapid review of the literature // *Education Sciences*. - 2023. - T. 13, № 4. - Article 410. <https://doi.org/10.3390/educsci13040410>
7. Zulpikhar Z., Rakhmetov M., Jugembayeva B., Kapanova D., Zhilmagambetova R. Assessment using artificial intelligence in higher education: Innovations and ethical challenges in the Ibero-American and Kazakh contexts: A mini-review // *Frontiers in Education*. - 2026. - T. 11. - Article 1798126. - <https://doi.org/10.3389/feduc.2026.1798126>
8. Abdrasilov B., Niyazov T., Shinetova L., Altybayeva S., Turalbayeva K., Orlov D. Generation of Kazakhstan's unified national testing variants using AI: A platform for automatic task creation with expert control // *Frontiers in Big Data*. - 2026. - T. 9. - Article 1772101. - <https://doi.org/10.3389/fdata.2026.1772101>
9. García-Varela F., Nussbaum M., Mendoza M., Martínez-Troncoso C., Bekerman Z. ChatGPT as a stable and fair tool for automated essay scoring // *Education Sciences*. - 2025. - T. 15, № 8. - Article 946. - <https://doi.org/10.3390/educsci15080946>
10. Seßler K., Fürstenberg M., Bühler B., Kasneci E. Can AI grade your essays? A comparative analysis of large language models and teacher ratings in multidimensional essay scoring // *Proceedings of the 15th International Learning Analytics and Knowledge Conference*. - New York : Association for Computing Machinery, 2025. - C. 462–472. - <https://doi.org/10.1145/3706468.3706527>
11. Karaçeper R. D., Kiray G. ChatGPT-4o as an automated scoring tool for writing assessment: Strengths and weaknesses // *International Journal of Assessment Tools in Education*. - 2026. - T. 13, № 1. - C. 66–94. - <https://doi.org/10.21449/ijate.1701871>
12. Wu H.-N., Chu M.-N., Hsu J.-L. Comparing GPT and human raters in essay assessment: Variability, bias, and the potential of LLM-based scoring // *Computers and Education Open*. - 2026. - T. 10. - Article 100341. - <https://doi.org/10.1016/j.cao.2026.100341>
13. Yermaganbetova M., Ashimbekova A., Kaibassova D., Akhitova R., Dyusseimbina E. Evaluating the impact of an AI-integrated learning platform on student performance: A quasi-experimental study // *Frontiers in Education*. - 2026. - T. 11. - Article 1792353. - <https://doi.org/10.3389/feduc.2026.1792353>
14. East M., Slomp D. The ethical turn in writing assessment: How far have we come, and where do we still need to go? // *Language Teaching*. - 2024. - T. 57, № 2. - C. 262–273. - <https://doi.org/10.1017/S0261444823000034>

Ә.Ү. Серікбаева, Д.М. Акижанова

ГЕНЕРАТИВТІ ЖАСАНДЫ ИНТЕЛЛЕКТ АДАМ ЭССЕСІН БАҒАЛАУШЫЛАРДЫ АЛМАСТЫРА АЛА МА? GPT-5.2, GEMINI ЖӘНЕ МҰҒАЛІМ БАҒАЛАУЛАРЫНЫҢ ПСИХОМЕТРИЯЛЫҚ САЛЫСТЫРМАЛЫ ТАЛДАУЫ

Оқушылардың жазба жұмыстарын бағалау маңызды, алайда көп уақытты қажет ететін үдеріс болып табылады. Үлкен тілдік модельдер (LLM) және генеративті жасанды интеллект саласындағы соңғы жетістіктер мұғалімдерге эсселерді бағалауда қолдау көрсетуге мүмкіндік беретін әлеуетті шешімдер ұсынады. Бұл зерттеуде GPT-5.2 және Gemini сияқты үлкен тілдік модельдер адам бағалаушыларымен салыстырылып, олардың сенімділігі психометриялық тұрғыдан бағаланды. Барлығы 55 эссе жинақталып, төрт критерийден тұратын бағалау рубрикасы негізінде GPT-5.2, Gemini және тәжірибелі мұғалім тарапынан бағаланды. Зерттеудің мақсаты – ағылшын тіліндегі эсселерді адам бағалаушыларының, GPT-5.2 және Gemini модельдерінің бағалау нәтижелерін психометриялық тұрғыдан салыстыру, олардың бағалау сенімділігін, өзара байланысын, бағалаудағы айырмашылықтарын және бағалау рубрикасының конструктілік валидтілігін анықтау. Нәтижелер GPT-5.2 сөздік қор критерийі бойынша ең жоғары ұпайларды, ал Gemini грамматика критерийі бойынша ең жоғары ұпайларды бергенін көрсетті. Жалпы алғанда, GPT-5.2 мұғалім бағалауларымен жоғары деңгейде сәйкестік көрсетті және эсселерді бағалауда сенімді қосымша құрал ретінде пайдаланылуы мүмкін. Дегенмен, күрделі мазмұн мен контекстік ерекшеліктерді бағалау кезінде мұғалімнің кәсіби пікірі маңызды болып қала береді. Алынған нәтижелер генеративті жасанды интеллектің ағылшын тілін шет тілі ретінде оқыту (EFL) контекстінде эсселерді бағалаудың тиімділігі мен бірізділігін арттыратын қосымша құрал ретінде пайдаланудың орынды екенін көрсетеді.

Түйін сөздер: үлкен тілдік модельдер (LLM), GPT-5.2, Gemini, эссені автоматтандырылған бағалау, адамдық бағалаушы, психометриялық бағалау, бағалаушылар арасындағы келісім

Ә.Ү. Серікбаева, Д.М. Акижанова

МОЖЕТ ЛИ ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ЗАМЕНИТЬ ЭКСПЕРТОВ-ОЦЕНЩИКОВ ЭССЕ? ПСИХОМЕТРИЧЕСКОЕ СРАВНЕНИЕ ОЦЕНОК GPT-5.2, GEMINI И ПРЕПОДАВАТЕЛЕЙ

Оценивание письменных работ учащихся является важным, но в то же время трудоемким процессом. Большие языковые модели (БЯМ) и другие современные достижения в области генеративного искусственного интеллекта предлагают потенциальные решения для оказания поддержки учителям при оценивании эссе. В данном исследовании проводится психометрическая оценка больших языковых моделей, таких как GPT-5.2 и Gemini, в сравнении с экспертами-оценщиками с целью определения надёжности данных решений. Всего было собрано 55 эссе, которые были оценены GPT-5.2, Gemini и опытным учителем с использованием оценочной рубрики, включающей четыре критерия. Цель данного исследования заключается в психометрическом сравнении результатов оценивания английских эссе экспертами-оценщиками, GPT-5.2 и Gemini, а также в определении надёжности оценивания, взаимосвязи между оценками, различий в оценивании и конструктивной валидности оценочной рубрики. Результаты показали, что GPT-5.2 присваивал наиболее высокие баллы по критерию «Лексика», в то время как Gemini демонстрировал наиболее высокие показатели по критерию «Грамматика». В целом GPT-5.2 показал высокий уровень согласованности с оценками учителя и может использоваться в качестве надёжного вспомогательного инструмента при оценивании эссе. Однако профессиональное суждение учителя остается важным при оценивании сложного содержания и контекстуальных особенностей. Полученные результаты подтверждают целесообразность использования генеративного искусственного интеллекта в качестве вспомогательного инструмента для повышения эффективности и согласованности оценивания эссе в контексте преподавания английского языка как иностранного (EFL).

Ключевые слова: автоматизированное оценивание эссе, большие языковые модели (БЯМ), GPT-5.2, Gemini, эксперт-оценщик, оценивание, психометрическая оценка.

References

1. Ma, M. (2023). Exploring learning-oriented assessment in EAP writing classrooms: Teacher and student perspectives. *Language Testing in Asia*, 13, Article 33. <https://doi.org/10.1186/s40468-023-00249-x>
2. Han, T., & Zheng, Y. (2026). AI and human scoring for postgraduate writing: Evaluating score reliability, variability, and rater behaviours. *Studies in Educational Evaluation*, 88, Article 101572. <https://doi.org/10.1016/j.stueduc.2026.101572>
3. Khamboonruang, A. (2023). Detecting differential rater severity in a high-stakes EFL classroom writing assessment: A many-facets Rasch measurement approach. *PASAA*, 66, 5–36.
4. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
5. Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), Article 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
6. Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), Article 410. <https://doi.org/10.3390/educsci13040410>
7. Zulpikhar, Z., Rakhmetov, M., Jugembayeva, B., Kapanova, D., & Zhilmagambetova, R. (2026). Assessment using artificial intelligence in higher education: Innovations and ethical challenges in the Ibero-American and Kazakh contexts: A mini-review. *Frontiers in Education*, 11, Article 1798126. <https://doi.org/10.3389/feduc.2026.1798126>
8. Abdrasilov, B., Niyazov, T., Shinetova, L., Altybayeva, S., Turalbayeva, K., & Orlov, D. (2026). Generation of Kazakhstan's unified national testing variants using AI: A platform for automatic task creation with expert control. *Frontiers in Big Data*, 9, Article 1772101. <https://doi.org/10.3389/fdata.2026.1772101>
9. García-Varela, F., Nussbaum, M., Mendoza, M., Martínez-Troncoso, C., & Bekerman, Z. (2025). ChatGPT as a stable and fair tool for automated essay scoring. *Education Sciences*, 15(8), Article 946. <https://doi.org/10.3390/educsci15080946>
10. Seßler, K., Fürstenberg, M., Bühler, B., & Kasneci, E. (2025). Can AI grade your essays? A comparative analysis of large language models and teacher ratings in multidimensional essay scoring. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 462–472). Association for Computing Machinery. <https://doi.org/10.1145/3706468.3706527>
11. Karaceper, R. D., & Kiray, G. (2026). ChatGPT-4o as an automated scoring tool for writing assessment: Strengths and weaknesses. *International Journal of Assessment Tools in Education*, 13(1), 66–94. <https://doi.org/10.21449/ijate.1701871>
12. Wu, H.-N., Chu, M.-N., & Hsu, J.-L. (2026). Comparing GPT and human raters in essay assessment: Variability, bias, and the potential of LLM-based scoring. *Computers and Education Open*, 10, Article 100341. <https://doi.org/10.1016/j.cao.2026.100341>
13. Yermaganbetova, M., Ashimbekova, A., Kaibassova, D., Akhitova, R., & Dyusseimbina, E. (2026). Evaluating the impact of an AI-integrated learning platform on student performance: A quasi-experimental study. *Frontiers in Education*, 11, Article 1792353. <https://doi.org/10.3389/feduc.2026.1792353>
14. East, M., & Slomp, D. (2024). The ethical turn in writing assessment: How far have we come, and where do we still need to go? *Language Teaching*, 57(2), 262–273. <https://doi.org/10.1017/S0261444823000034>

APPENDICES**Appendix A. Student Essay Writing Task****Essay Topic: The Role of Education in Shaping a Successful Future****Instructions**

- Word limit: **150–180 words**
- Introduction (30–40 words)
- Body paragraph 1 (45–55 words)
- Body paragraph 2 (45–55 words)
- Conclusion (20–30 words)

Prompt Used for GPT-5.2 and Gemini

The following prompt was used identically for GPT-5.2 and Gemini to ensure consistency across AI-based evaluations.

You are an experienced EFL writing examiner. Evaluate the following student essay using the analytic writing rubric below.

Criteria:

1. Organization (1–5)
2. Vocabulary (1–5)
3. Grammar (1–5)
4. Task Achievement (1–5)

Instructions:

- Read the essay carefully.
- Assign a score from 1 to 5 for each criterion.
- Calculate the total score out of 20.
- Provide brief rubric-based feedback for each criterion.
- Do not rewrite or correct the essay.
- Evaluate each essay independently without considering previous responses.

Appendix B. Example of Teacher Assessment

Criterion	Score
Organization	3.5
Vocabulary	3.5
Grammar	3.0
Task Achievement	4.5
Total	14.5/20

Appendix C. Example of GPT-5.2 Evaluation

Criterion	Score	Feedback
Organization	4.0	Clear overall structure with an introduction, body paragraphs, and a conclusion. Some transitions could be improved.
Vocabulary	3.5	Appropriate vocabulary, although several collocations are unnatural and some repetition occurs.
Grammar	3.0	Frequent grammatical errors, particularly subject–verb agreement and article usage, reduce accuracy.
Task Achievement	3.5	The topic is addressed, but the essay is below the required word limit and ideas could be developed further.
Total	14.0/20	

Appendix D. Example of Gemini Evaluation

Criterion	Score	Feedback
Organization	3.5	The essay follows the required structure, but paragraphing and logical flow can be strengthened.
Vocabulary	3.5	A sufficient range of vocabulary is demonstrated, although spelling and word-choice errors are present.
Grammar	3.0	Several grammatical inaccuracies affect fluency and overall language accuracy.
Task Achievement	4.5	The response addresses the task effectively with relevant ideas and an appropriate example, although the conclusion is rather brief.
Total	14.5/20	

Авторлар туралы мәліметтер:

Серікбаева Әсия Үмбетбайқызы (корреспондент автор) - Әл-Фараби атындағы Қазақ ұлттық университеті, Түркітану және тіл теориясы кафедрасының 3-курс докторанты, Қазақстан, Алматы қаласы e-mail: assiyaserikbayeva237@gmail.com

Акижанова Диана Муратовна, PhD докторы, Л.Н.Гумилев атындағы Еуразия ұлттық университетінің «Аударма теориясы мен практикасы» кафедрасының меңгерушісі, Қазақстан, Астана қаласы e-mail: akizhanova_dm@enu.kz

Сведения об авторах:

Серікбаева Әсия Үмбетбаевна (автор-корреспондент) - докторант 3 курса кафедры тюркологии и теории языка Казахского национального университета имени аль-Фараби, Республика Казахстан, Алматы, e-mail: assiyaserikbayeva237@gmail.com

Акижанова Диана Муратовна - доктор PhD, заведующая кафедрой «Теория и практика перевода» Евразийского национального университета имени Л.Н. Гумилева. Республика Казахстан, Астана, e-mail: akizhanova_dm@enu.kz

Information about the authors:

Serikbayeva Assiya Umbetbaykyzy (corresponding author) - 3rd-year PhD Candidate, Department of Turkology and Language Theory, Al-Farabi Kazakh National University Kazakhstan, Almaty, e-mail: assiyaserikbayeva237@gmail.com

Akizhanova Diana Muratovna - PhD Head of the Department of Translation Theory and Practice, L.N. Gumilyov Eurasian National University, Kazakhstan, Astana, e-mail: akizhanova_dm@enu.kz