Information about the authors

Gulsipat Abisheva – doctoral student of the Department of artificial intelligence, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan; e-mail: gulsipat.abisheva@gmail.com. ORCID: https://orcid.org/0009-0003-6692-197X.

Bibigul Razakhova – c.t.s., Head of the Department of Artificial Intelligence Technologies, L.N. Gumilev Eurasian National University, Astana, Kazakhstan; e-mail: razakhova_bsh@enu.kz. ORCID: https://orcid.org/0000-0002-8152-8661.

Tolegen Aidynov* – doctoral student of the Department of Information Security, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan; e-mail: tolegen.ch@gmail.com. ORCID: http://orcid.org/0009-0005-3327-9719.

Minira Saparbek – master of the Department of artificial intelligence, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan; e-mail: sapminira@gmail.com. ORCID: https://orcid.org/0009-0005-0334-0910.

Поступила в редакцию 19.07.2026
Поступила после доработки 29.08.2026
Принята к публикации 01.09.2026

https://doi.org/10.53360/2788-7995-2026-3(23)-23

CC BY 4.0
Check for updates

IRSTI: 28.23.02

K.E. Iklassova[1]*, L.N. Abdullah[2], M.Zh. Bazarova[3]
[1]M. Kozybayev North Kazakhstan University,
150000, Republic of Kazakhstan, Petropavllovsk, Pushkina Str, 86
[2]Putra University of Malaysia, Malaysia, Serang, Selangor, 43400 Seri Kembangan
[3]Sarsen Amanzholov East Kazakhstan University,
070002, Republic of Kazakhstan, Ust-Kamenogorsk, 30th Gvardeiskoy Divisii Str, 34
*e-mail: keiklasova@ku.edu.kz

ASSESSMENT OF INSTITUTIONAL DATA HETEROGENEITY FOR ARTIFICIAL INTELLIGENCE SYSTEMS

Abstract: *This study quantifies institutional heterogeneity in digital engagement and academic outcomes using anonymized data from Kazakhstani schools, comprising records from two EdTech platforms (Twig-bilim.kz, BilimLand.kz) and entry diagnostic test scores. Applying one-way ANOVA/intraclass correlation (ICC) decomposition and random-intercept mixed-effects modeling, we find that school membership explains 8.8-9.8% of variance in digital platform engagement and 30.0-30.7% of variance in entry test scores, with close numerical agreement between the ANOVA-based variance shares and the mixed-effects ICC estimates. These estimates indicate non-negligible school-level heterogeneity, particularly for entry diagnostic-test scores. The findings indicate that pooled, school-agnostic machine learning models risk systematically miscalibrated predictions for students in schools that deviate from the national average, and argue for school-aware calibration in the design of AI-based early-warning and adaptive-learning systems intended to support flexible and equitable access to education. The results are discussed in relation to Kazakhstan's ongoing school digitalization initiatives and to the broader literature on algorithmic fairness in education.*

Key words: *artificial intelligence, learning analytics, institutional heterogeneity, intraclass correlation, multilevel modeling, algorithmic fairness, educational data mining.*

Introduction

The development of artificial intelligence (AI) in education – from early warning systems to adaptive personalized-learning platforms – is increasingly regarded as an important means of improving the flexibility and accessibility of education. Recent systematic reviews indicate that generative AI, intelligent tutoring systems, and learning analytics are increasingly positioned as technologies capable of transforming personalized learning from a static, uniform model into an adaptive, data-driven process that supports dynamic learning pathways [1].

©K.E. Iklassova, L.N. Abdullah, M.Zh. Bazarova, 2026

At the same time, Kazakhstan – a country in which education digitalization is a national prioritycontinues to face disparities in digital infrastructure between urban and rural schools. Empirical evidence from a nationally representative teacher sample shows that the digital-access divide narrows when schools provide teachers with ICT resources; internet quality and rural-urban location are also significantly associated with the use of synchronous digital tools [2]. More recent public reporting continues to highlight uneven infrastructure and digital readiness, raising the risk that AI deployment may reproduce rather than reduce existing educational inequalities [3].

This creates a methodological issue for AI-based learning-support systems: when educational data are analyzed using pooled models without explicitly accounting for school membership, their hierarchical structure may be overlooked. Students are nested within schools that may differ in digital readiness and other contextual characteristics. Multilevel modeling explicitly represents this institutional level, and the intraclass correlation coefficient (ICC) quantifies the proportion of outcome variance attributable to between-cluster differences. Ignoring clustering can distort standard errors and statistical inference for individual-level associations [4].

Study [5] applies school-level random-intercept multilevel models to examine the relationship between institutional AI readiness and student AI literacy, using an unconditional multilevel model to quantify between-institution variance before adding higher-level predictors. This supports the methodological relevance of multilevel analysis in contemporary AIED research. However, such work has largely focused on large national survey or linked administrative datasets rather than behavioral logs from EdTech platforms in an emerging digital ecosystem.

A systematic mapping study of machine-learning fairness in education identifies a growing but fragmented research landscape in which unfairness may arise both from pre-existing structural inequalities and from algorithmic decisions that reproduce or amplify them [6]. Particularly relevant is [7], which uses the Open University Learning Analytics Dataset (OULAD) and reports institutional data as the strongest contributor to measured model discrimination among the examined data-source categories, followed by VLE activity data, with assessment data contributing less. This provides direct empirical evidence that institutional information can be a material source of systematic bias in educational AI models.

A recent study [8] directly audits intersectional fairness across gender and rural–urban location in an academic early warning system. The study demonstrates that aggregate performance can conceal substantial subgroup disparities and that fairness post-processing changes error trade-offs across intersectional groups. It is a relevant example of explicitly testing geographic and demographic inequality in an EWS; however, it is based on a single Moroccan region and does not establish whether comparable institutional effects generalize to other national settings, including post-Soviet Central Asia.

The scoping review of adaptive mechanisms for learning management systems [9] synthesizes adaptive-learning approaches and shows that the literature is predominantly learner-centered, with adaptation driven mainly by individual characteristics, behavior, or performance. Institutional context is rarely treated as an explicit clustering level in the design or evaluation of adaptive mechanisms. This supports the gap addressed here: school or institution membership is seldom modeled explicitly when assessing the performance or potential bias of adaptive systems.

The updated systematic review of personalized learning in the AI era [1] confirms that personalization is increasingly conceptualized as a flexible, data-driven process. Nevertheless, the literature remains centered primarily on learner-level personalization, terminology, privacy, bias, and ethical issues rather than on heterogeneity in the institutional environments in which AI-enabled learning systems are deployed.

Studies of digitalization in Kazakhstan’s higher education sector report persistent institutional barriers and uneven readiness despite expanded access to digital technologies [10]. Evidence from secondary education during pandemic-era online learning similarly documents rural–urban and regional disparities in access, experience, and technological conditions [11]. These studies, however, rely primarily on survey or self-reported data rather than behavioral logs from EdTech platforms and do not quantify the magnitude of school-level variance using multilevel variance decomposition. The scale of the institutional effect in operational educational data therefore remains insufficiently established.

In 2026, Kazakhstan announced the Qazaq Digital Mektebi pilot, intended to reduce disparities in educational quality between rural and urban schools through the use of AI-supported teaching

approaches [12]. Thus, policy decisions concerning the scaling of AI in schools are already being made, while open empirical evidence quantifying institutional heterogeneity for the calibration and evaluation of such systems remains limited.

The reviewed literature points to three intersecting gaps relevant to the present study. First, multilevel variance decomposition and ICC estimation are not consistently incorporated as diagnostic tools when the suitability of educational data for pooled AI models is assessed. Second, educational algorithmic-fairness studies more commonly examine individual-level attributes than institutional or geospatial heterogeneity. Third, the cited literature on Kazakhstan documents digital inequalities primarily through survey-based or descriptive evidence, while quantitative estimates of school-level heterogeneity based on behavioral EdTech logs remain limited. Against this background, the present study estimates the magnitude of between-school heterogeneity and examines its implications for AI-system design.

To quantify the magnitude of institutional (between-school) heterogeneity in digital engagement and learning outcomes using operational data from Kazakhstani schools and to assess its relevance to the calibration and validation of AI-based learning-support systems. Research Objectives: 1) To conduct descriptive and statistical analyses of between-school variation in students' digital activity on educational platforms. 2) To analyze between-school variation in entry diagnostic-test results and compare the magnitude of the institutional effect with that observed for digital activity. 3) To estimate school-level variance components using random-intercept mixed-effects models and compare them with ANOVA-based variance shares. 4) To formulate practical implications of the identified institutional heterogeneity for the design of early warning and adaptive-learning AI systems. 5) To relate the findings to the broader context of the digital divide in Kazakhstan and current school-digitalization initiatives.

Research Methods

The empirical dataset is an anonymized institutional extract of user activity and student test results from Kazakhstani schools. It contains three worksheets: (1) "BC tests," comprising entry diagnostic-test results for Grades 3-4 and functional-literacy test results; (2) "Twig," containing counts of student visits to Twig-bilim.kz; and (3) "BilimLand," containing analogous activity data for BilimLand.kz. Records are identified only by anonymized school_id and user_id codes; the extract contains no directly identifying personal data. A key limitation is that the school_id sets in the testing worksheet and the platform-activity worksheets do not overlap (zero shared school_id values). Digital-activity and academic-outcome measures therefore refer to different school subsamples and are analyzed as separate case studies (Table I).

Table 1 – Structure of the analyzed data subsamples

Subsample	N observations	N students	N schools	Outcome
Digital activity (Twig+BilimLand)	1784	1784	22	Total visits
Entry diagnostic tests	743	495	52	Score, %
Functional literacy (illustrative)	257	95	3	Score, %

The analysis was conducted in Python 3 using Google Colab, with the source file uploaded by the analyst. Tabular data processing used pandas [13]; statistical testing used scipy.stats; mixed-effects modeling used statsmodels [14]; and visualization used matplotlib and seaborn. The code was organized into reproducible modules.

Because the distribution of platform-visit counts was strongly right-skewed, a logarithmic transformation was used for statistical analysis of activity:

$$y' = \ln(1 + visits) \quad (1)$$

Between-school heterogeneity was quantified using two complementary approaches.

(a) One-way ANOVA with the proportion of total sum of squares attributable to between-school differences (η^2):

$$\eta^2 = SSbetween / (SSbetween + SSwithin) \quad (2)$$

(b) A school-level random-intercept mixed-effects model estimated by restricted maximum likelihood (REML):

$$Y_{ij} = \beta_0 + u_j + \varepsilon_{ij} \quad (3)$$
$$u_j \sim N(0, \sigma^2_b), \ \varepsilon_{ij} \sim N(0, \sigma^2_w)$$

where Yij is the outcome for student i in school j; β0 is the overall fixed intercept; uj is the random school effect; and εij is the residual error. The model-based ICC is:

$$ICC = \sigma^2_b / (\sigma^2_b + \sigma^2_w) \quad (4)$$

1) The study is cross-sectional and does not support causal inference. 2) Platform activity and test outcomes come from non-overlapping school subsamples. 3) The functional-literacy subsample (three schools) is illustrative only. 4) Schools with n < 3 were excluded from comparisons of school mean scores. 5) The extract is not guaranteed to constitute a nationally representative sample of Kazakhstani schools.

Results

Analysis of the combined Twig-bilim.kz and BilimLand.kz data (N = 1,784 students, 22 schools) yielded a mean of 24.64 visits (SD = 67.29) with pronounced right skew. School-level means ranged from 7.2 to 86.6 visits, a twelvefold difference. To examine the between-school structure of digital engagement in greater detail, the distribution of log-transformed platform activity was compared across the 22 schools. As shown in Fig. 1, both the central tendency and dispersion of student activity vary across schools. The logarithmic scale reduces the influence of the strongly right-skewed visit distribution while preserving visible differences in school-level activity patterns.

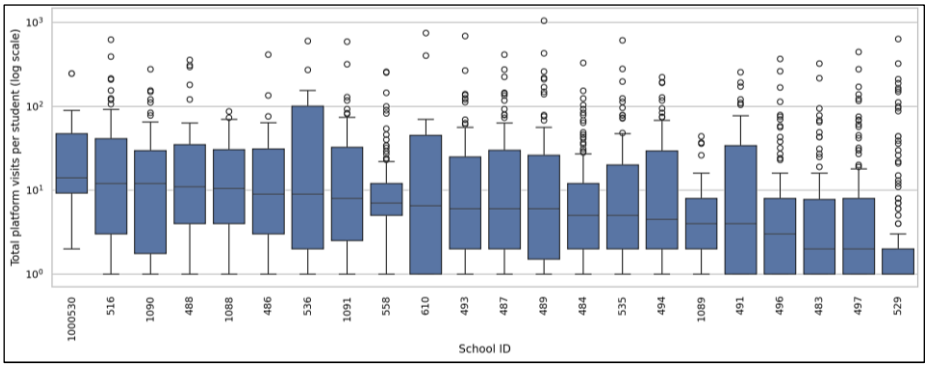


Fig. 1 – Distribution of student digital activity across schools (log scale)

To complement the distributional comparison, school-level mean activity was estimated together with 95% confidence intervals. Fig. 2 shows substantial variation in the average number of platform visits across schools, with school means ranging from 7.2 to 86.6 visits. The confidence intervals also reflect differences in within-school variability and sample size, providing a descriptive representation of the institutional heterogeneity subsequently quantified by the ANOVA-based variance share and the mixed-effects model.

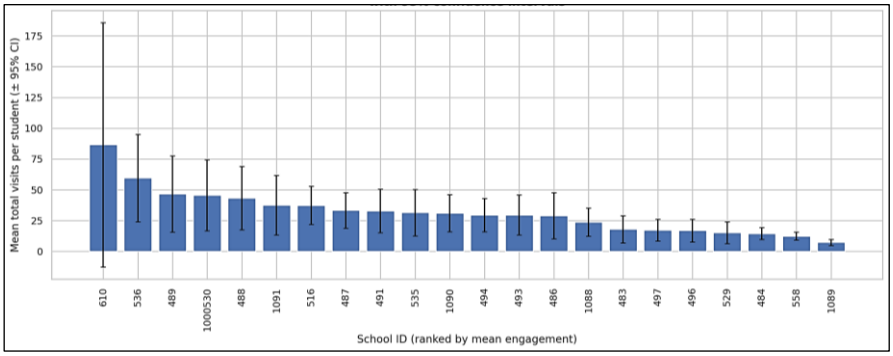


Fig. 2 – Mean number of visits by school with 95% confidence intervals

ANOVA on the log-transformed variable yielded F = 9.17, p = 5.75 × 10^{-28}, with an ANOVA-based between-school variance share η² = 0.098. The mixed-effects model yielded σ²b = 0.153 and σ²w = 1.597, corresponding to ICC = 0.088. The two complementary measures therefore indicate a similar magnitude of school-level heterogeneity.

Analysis of entry diagnostic tests (N = 743 records, 495 students, 52 schools) revealed substantially greater between-school variation. The overall mean score was 79.0%, while school means ranged from 26.0% to 95.0%, a gap of 69 percentage points. The between-school variation

in academic outcomes is further illustrated in Fig. 3, which presents the distribution of entry diagnostic-test scores for schools with at least three observations. The figure reveals marked differences in both school-level central tendencies and within-school score distributions, indicating that the overall mean of 79.0% masks considerable institutional variation

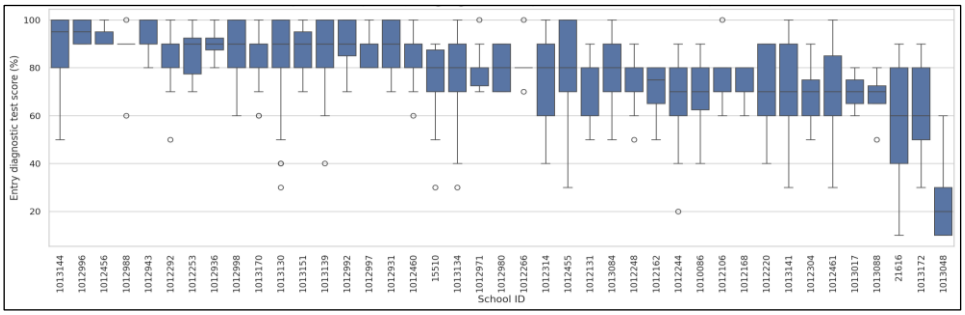


Fig. 3 – Entry diagnostic-test scores by school (n ≥ 3)

Because estimates of school-level means may be sensitive to the number of tested students, the relationship between school sample size and mean diagnostic-test performance was examined separately. As shown in Fig. 4, smaller schools exhibit greater dispersion in their mean scores, forming a visible funnel-like pattern. However, deviations from the overall mean are also observed among several schools with larger samples, suggesting that differences in sample size alone are unlikely to account for all of the observed between-school variation.

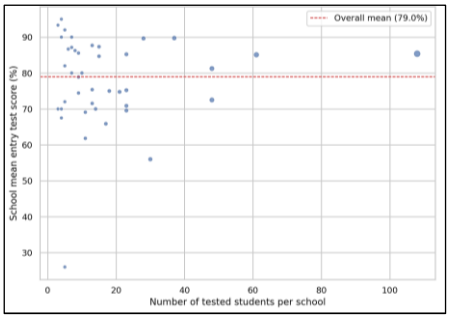


Fig. 4 – School mean score versus number of tested students

ANOVA yielded $F = 6.76$, $p = 5.97 \times 10^{-32}$, with $\eta^2 = 0.307$. The mixed-effects model yielded $\sigma^2 b = 99.71$ and $\sigma^2 w = 232.80$, corresponding to ICC = 0.300. Schools with $n < 10$ show wider dispersion, consistent with a funnel pattern; however, several larger schools ($n > 60$) also remain visibly displaced from the overall mean, indicating that the observed between-school differences cannot be attributed solely to small-sample variability.

The functional-literacy subsample (N = 257 records, 95 students, 3 schools) yielded an ANOVA-based between-school variance share $\eta^2 = 0.122$ ($F = 17.69$; $p = 6.43 \times 10^{-8}$). Although the functional-literacy subsample contains only three schools and is therefore insufficient for generalizable inference, it provides an additional descriptive view of institutional variation across assessment domains. Fig. 5 presents functional-literacy scores by school and domain, illustrating differences in score distributions across the three institutions. Given the limited number of schools, this result is treated as illustrative and is not used as primary evidence for the study's conclusions.

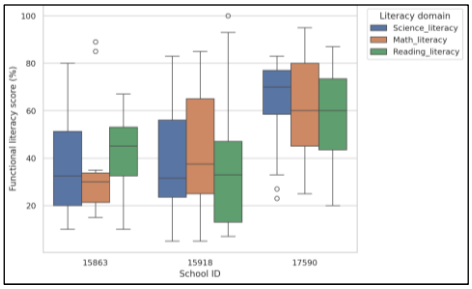


Fig. 5 – Functional-literacy scores by school and domain (three schools)

To compare the magnitude of between-school heterogeneity across the analyzed outcomes, the estimated variance measures are summarized in Fig. 6. The strongest institutional clustering is observed for entry diagnostic-test scores, for which the ANOVA-based between-school variance share reaches $\eta^2 = 0.307$ and the mixed-effects ICC is 0.300. Digital activity shows a smaller but non-negligible between-school component ($\eta^2 = 0.098$; ICC = 0.088), whereas the functional-literacy estimate ($\eta^2 = 0.122$) should be interpreted cautiously because it is based on only three schools.

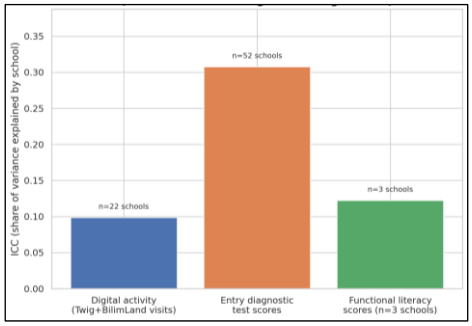


Fig. 6 – Comparison of between-school variance measures across the three outcomes

The numerical results underlying this comparison are summarized in Table 2. For both primary outcomes, the ANOVA-based between-school variance shares and the ICC estimates obtained from the random-intercept models are numerically close: $\eta^2 = 0.098$ versus ICC = 0.088 for digital activity, and $\eta^2 = 0.307$ versus ICC = 0.300 for entry diagnostic-test scores. Although η^2 and ICC are not identical statistical estimands, their close numerical agreement provides convergent evidence of between-school heterogeneity in both analyzed subsamples.

Table 2 – Summary statistical results. $H^2(a)$ – anova-based variance share; icc(m) – mixed-effects icc. * – illustrative subsample

Outcome	N schools	$\eta^2(A)$	ICC(M)	F	p
Digital activity	22	0,098	0,088	9,17	5,75e-28
Entry tests	52	0,307	0,300	6,76	5,97e-32
Functional literacy*	3	0,122	—	17,69	6,43e-8

The functional-literacy result is reported separately as an illustrative estimate because the corresponding subsample includes only three schools.

Discussion

The results provide a quantitative answer to the question posed in the Introduction: institutional heterogeneity in these Kazakhstani educational data is statistically significant and substantively non-negligible. Hedges and Hedberg [15] report substantial variation in school-level ICCs across educational outcomes and grade levels, supporting the need to account for clustering in educational research. In the present data, the mixed-model ICC is 0.088 for digital activity and 0.300 for entry diagnostic-test scores, indicating considerably stronger school-level clustering for academic outcomes than for platform engagement.

The close numerical agreement between the ANOVA-based variance shares ($\eta^2 = 0.098$ and 0.307) and the mixed-model ICCs (0.088 and 0.300) provides convergent evidence of between-school heterogeneity. Because η^2 and ICC are related but not identical estimands, their agreement is interpreted as complementary evidence rather than as replication by two independent ICC estimators.

The observed institutional effect is consistent with the broader literature on digital inequality in Kazakhstan at both higher-education [10] and secondary-school levels [2], [11]. The present study extends that literature by expressing institutional heterogeneity as a quantitative variance component in operational educational data. From an AI-design perspective, the findings are also consistent with [7], which identifies institutional data as an important source of measured algorithmic bias in OULAD-based student-progress models, and with [8], which shows that EWS error rates can differ across rural–urban and gender intersections. Because the present samples are not nationally representative and are not linked across outcomes, the magnitude observed here should not be interpreted as a national benchmark.

The practical implication is that school membership should be considered explicitly during model development and validation. A non-zero school-level ICC indicates statistical dependence among observations within schools and provides a reason to test whether pooled models generalize and calibrate consistently across institutions. The present analysis does not train or compare pooled and school-aware predictive models; therefore, any improvement in predictive accuracy or calibration from school-aware modeling remains to be established in subsequent validation studies.

The non-overlapping subsamples prevented testing the student-level relationship between platform activity and academic outcomes. The functional-literacy subsample includes only three schools and cannot support generalization. The study is cross-sectional and does not permit causal inference. The institutional extract is not known to be nationally representative. In addition, the study quantifies clustering but does not directly evaluate predictive-model calibration or fairness metrics across schools.

Conclusion

The analysis confirms substantial institutional heterogeneity in the examined Kazakhstani educational data. For digital engagement, the ANOVA-based between-school variance share is 0.098 and the mixed-model ICC is 0.088; for entry diagnostic-test scores, the corresponding values are 0.307 and 0.300. Thus, school-level clustering is markedly stronger for academic outcomes than for platform engagement, and the magnitude is similar across the two complementary variance-decomposition approaches.

These findings support treating institutional clustering as a design and validation consideration for AI-based learning-support systems in the studied Kazakhstani context, rather than assuming a priori that a single pooled model will perform uniformly across schools. Given ongoing school-digitalization initiatives, including the announced Qazaq Digital Mektebi pilot [12], school-level variance diagnostics may be incorporated into model-development and evaluation workflows before large-scale deployment.

Future research should: 1) obtain school geolocation or rural–urban indicators to test geospatial-fairness hypotheses; 2) collect linked platform-activity and outcome data for the same students and schools; 3) validate the estimated ICCs in a nationally representative sample; and 4) directly compare pooled and school-aware predictive models using calibration and fairness metrics.

References

1. Khalifeh F. Redefining personalized learning in the artificial intelligence era: An updated systematic review from 2019 to 2025 / F. Khalifeh, R. Santiago, R. Palau // Smart Learning Environments. – 2026. – Vol. 13. – Art. 19. https://doi.org/10.1186/s40561-026-00440-6.
2. The impact of the digital divide on synchronous online teaching in Kazakhstan during COVID-19 school closures / A. Amirova et al // Frontiers in Education. – 2023. – Vol. 7. – Art. 1083651. https://doi.org/10.3389/feduc.2022.1083651.
3. How artificial intelligence transforms education in Kazakhstan // The Astana Times. – 2026. – 7 January. https://astanatimes.com/2026/01/how-artificial-intelligence-transforms-education-in-kazakhstan/ (date of access: 02.09.2026).
4. Raudenbush S.W. Hierarchical Linear Models: Applications and Data Analysis Methods / S.W. Raudenbush, A.S. Bryk. – 2nd ed. – Thousand Oaks, CA: Sage Publications, 2002.
5. From school AI readiness to student AI literacy: A national multilevel mediation analysis of institutional capacity and teacher capability / X. Guan et al // arXiv. – 2026. – arXiv:2603.20056. – https://doi.org/10.48550/arXiv.2603.20056.
6. To N.P. Fairness for machine learning software in education: A systematic mapping study / N.P. To, P.N. Hung, A. NguyenDuc // SSRN. – 2024. https://doi.org/10.2139/ssrn.4713827.
7. Idowu J.A. Investigating algorithmic bias in student progress monitoring / J.A. Idowu, A.S. Koshiyama, P. Treleaven // Computers and Education: Artificial Intelligence. – 2024. – Vol. 7. – Art. 100267. https://doi.org/10.1016/j.caeai.2024.100267.
8. El Jihaoui M. Towards a fair and transparent early warning system for identifying students at risk of academic failure / M. El Jihaoui, O.E.K. Abra, K. Mansouri // Frontiers in Education. – 2026. – Vol. 11. – Art. 1878686. https://doi.org/10.3389/feduc.2026.1878686.
9. Adaptive learning mechanisms for learning management systems: A scoping review and practical considerations / S. Kucharski et al // arXiv. – 2025. – arXiv:2512.18383.
10. The role of digital technologies in higher education institutions: The case of Kazakhstan / D. Nurtayeva et al // Problems and Perspectives in Management. – 2024. – Vol. 22, № 1. – P. 562-577. https://doi.org/10.21511/ppm.22(1).2024.45.

11. COVID-19 online learning challenges: Kazakhstan secondary schools case study / S. Nurgaliyeva et al // Frontiers in Education. – 2024. – Vol. 9. – Art. 1448594. https://doi.org/10.3389/feduc.2024.1448594.
12. Pokidaev D. Kazakhstan to launch digital schools project / D. Pokidaev // The Times of Central Asia. – 2026. – 2 April. https://timesca.com/kazakhstan-to-launch-digital-schools-project/ (date of access: 02.09.2026).
13. McKinney W. Data structures for statistical computing in Python / W. McKinney // Proceedings of the 9th Python in Science Conference (SciPy 2010). – 2010. – P. 56-61. https://doi.org/10.25080/Majora-92bf1922-00a.
14. Seabold S. Statsmodels: Econometric and statistical modeling with Python / S. Seabold, J. Perktold // Proceedings of the 9th Python in Science Conference (SciPy 2010). – 2010. – P. 92-96. https://doi.org/10.25080/Majora-92bf1922-011.
15. Hedges L.V. Intraclass correlation values for planning group-randomized trials in education / L.V. Hedges, E.C. Hedberg // Educational Evaluation and Policy Analysis. – 2007. – Vol. 29, № 1. – P. 60-87. https://doi.org/10.3102/0162373707299706.

Funding

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant № AP23488869)

К.Е. Икласова[1*], Л.Н. Абдулла[2], М.Ж. Базарова[3]
[1]Северо-Казахстанский университет имени М. Козыбаева,
150000, Республика Казахстан, г. Петропавловск, ул. Пушкина, 86
[2]Университет Путра Малайзия, Малайзия, Серданг, Селангор, 43400 Сери Кембанган
[3]Восточно-Казахстанский университет имени С. Аманжолова,
070002, Республика Казахстан, г. Усть-Каменогорск, ул. 30-й Гвардейской дивизии, 34
*e-mail: keiklasova@ku.edu.kz

ОЦЕНКА ИНСТИТУЦИОНАЛЬНОЙ ГЕТЕРОГЕННОСТИ ДАННЫХ ДЛЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В работе количественно оценивается институциональная гетерогенность цифровой вовлечённости и учебных результатов на анонимизированных данных школ Казахстана, включающих записи двух EdTech-платформ (Twig-bilim.kz, BilimLand.kz) и баллы входного диагностического тестирования. С помощью однофакторного ANOVA/декомпозиции коэффициента внутриклассовой корреляции (ICC) и многоуровневого моделирования со случайным пересечением показано, что принадлежность к школе объясняет 8,8-9,8% дисперсии цифровой вовлечённости и 30,0-30,7% дисперсии баллов входного тестирования, при этом ANOVA-оценки доли межшкольной дисперсии и ICC из модели смешанного влияния численно близки. Полученные оценки указывают на содержательно заметную межшкольную гетерогенность, особенно для результатов входного диагностического тестирования. Результаты показывают, что институциональную кластеризацию следует явно учитывать при оценке калибровки и обобщающей способности моделей машинного обучения между школами с различающимися распределениями результатов, и обосновывают необходимость валидации при проектировании ИИ-систем раннего предупреждения и адаптивного обучения, призванных поддерживать гибкий и справедливый доступ к образованию. Результаты обсуждаются в контексте текущих инициатив цифровизации школ Казахстана и более широкой литературы об алгоритмической справедливости в образовании.

Ключевые слова: *искусственный интеллект, аналитика обучения, институциональная гетерогенность, коэффициент внутриклассовой корреляции, многоуровневое моделирование, алгоритмическая справедливость, интеллектуальный анализ образовательных данных.*

К.Е. Икласова[1*], Л.Н. Абдулла[2], М.Ж. Базарова[3]
[1]М.Қозыбаев атындағы Солтүстік Қазақстан университеті,
150000, Қазақстан Республикасы, Петропав қаласы, Пушкин көшесі, 86
[2]Putra университеті, Малайзия, Серанг, Селангор, 43400 Сери Кембанган
[3]С. Аманжолов атындағы Шығыс Қазақстан университеті,
070002, Қазақстан Республикасы, Өскемен қаласы, 30-шы Гвардиялық дивизия, 34 к.
*e-mail: keiklasova@ku.edu.kz

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖҮЙЕЛЕРІНЕ АРНАЛҒАН ДЕРЕКТЕРДІҢ ИНСТИТУЦИОНАЛДЫҚ ГЕТЕРОГЕНДІЛІГІН БАҒАЛАУ

Бұл жұмыста Қазақстан мектептерінің анонимделген деректері негізінде цифрлық қатысу мен оқу нәтижелеріндегі институционалдық (мектепаралық) гетерогенділік сандық түрде бағаланады; деректер екі EdTech-платформасының (Twig-bilim.kz, BilimLand.kz) жазбаларын және кіру диагностикалық тестілеу балдарын қамтиды. Бірфакторлы ANOVA/ішкі сыныптық корреляция коэффициентін (ICC) ыдырату және кездейсоқ қиылысы бар көпдеңгейлі модельдеу арқылы мектепке тиесілілік цифрлық қатысу дисперсиясының 8,8-9,8%-ын және кіру тестілеу балдары дисперсиясының 30,0-30,7%-ын түсіндіретіні анықталды, әрі ANOVA негізіндегі мектепаралық дисперсия үлесі мен mixed-effects моделіндегі ICC сандық тұрғыдан жақын болды. Алынған бағалар мектепаралық гетерогенділіктің, әсіресе кіру диагностикалық тестілеу нәтижелері үшін, мазмұндық тұрғыдан елеулі екенін көрсетеді. Нәтижелер нәтижелерінің үлестірімі әртүрлі мектептер арасында машиналық оқыту модельдерінің калибрленуі мен жалпылану қабілетін бағалау кезінде институционалдық кластерленуді айқын ескеру қажеттігін көрсетеді және білімге икемді әрі әділ қолжетімділікті қолдауға арналған ерте ескерту мен бейімделген оқыту ЖИ-жүйелерін жобалауда мектепті ескеретін валидацияның маңызын негіздейді. Нәтижелер Қазақстандағы мектептерді цифрландыру бойынша ағымдағы бастамалар және білім берудегі алгоритмдік әділдік туралы кеңірек әдебиет аясында талқыланады.

Түйінді сөздер: *жасанды интеллект, оқыту аналитикасы, институционалдық гетерогенділік, ішкі сыныптық корреляция коэффициенті, көпдеңгейлі модельдеу, алгоритмдік әділдік, білім беру деректерін өндіру.*

Сведения об авторах

Кайнижамал Есимсеитовна Икласова* – PhD, доцент кафедры «Информационно-коммуникационные технологии», Северо-Казахстанский университет имени М. Козыбаева; e-mail: keiklasova@ku.edu.kz. ORCID: https://orcid.org/0000-0002-8330-4282.

Лили Нурлияна Абдулла – PhD, ассоциированный профессор кафедры мультимедиа, факультет компьютерных наук и информационных технологий, Университет Путра Малайзия; e-mail: liyana@upm.edu.my. ORCID: https://orcid.org/0000-0001-8704-2390.

Мадина Жомартовна Базарова – PhD, ассоциированный профессор кафедры «Компьютерное моделирование и информационные технологии», Восточно-Казахстанский университет имени С. Аманжолова; e-mail: madina_vkgtu@mail.ru. ORCID: https://orcid.org/0000-0003-2580-6580.

Авторлар туралы мәліметтер

Кайнижамал Есимсеитовна Икласова* – PhD, «Ақпараттық-коммуникациялық технологиялар» кафедраның доценті, М. Козыбаев атындағы Солтүстік-Қазақстан университеті; e-mail: keiklasova@ku.edu.kz. ORCID: https://orcid.org/0000-0002-8330-4282.

Лили Нурлияна Абдулла – PhD, мультимедиа кафедрасының қауымдастырылған профессоры, Компьютерлік ғылымдар және ақпараттық технологиялар факультеті, Путра Малайзия университеті; e-mail: liyana@upm.edu.my. ORCID: https://orcid.org/0000-0001-8704-2390.

Мадина Жомартовна Базарова – PhD, «Компьютерлік үлгілеу және ақпараттық технологиялар» кафедрасының қауымдастырылған профессоры, С. Аманжолов атындағы Шығыс Қазақстан университеті; e-mail: madina_vkgtu@mail.ru. ORCID: https://orcid.org/0000-0003-2580-6580.

Information about the authors

Kainizhamal Iklassova* – PhD, associate professor, Department of Information and Communication Technologies, Manash Kozybayev North Kazakhstan University; e-mail: keiklasova@ku.edu.kz. ORCID: https://orcid.org/0000-0002-8330-4282.

Lili Nurliyana Abdullah – PhD, Associate Professor, Department of Mulitimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia; e-mail: liyana@upm.edu.my. ORCID: https://orcid.org/0000-0001-8704-2390.

Madina Zhomartovna Bazarova – PhD, associate professor of the Department of Computer Modeling and Information Technology, Sarsen Amanzholov East Kazakhstan University; e-mail: madina_vkgtu@mail.ru. ORCID: https://orcid.org/0000-0003-2580-6580.

Received 16.06.2026
Revised 18.08.2026
Accepted 31.08.2026