

¹Zhanar Bekbayeva, ²Moldir Seidina, ³Saule Aldabergenova, ⁴Botagoz Bermukhambetova*,
⁵Pawel Bogacz

¹ L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan

² Kazakh National University of Sports (KNUS), Astana, Republic of Kazakhstan

³ Saken Seifullin Kazakh Agro Technical Research University (KazATU), Astana, Republic of Kazakhstan

⁴ National Center for Development Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan

⁴ AGH University of Krakow, Krakow, Republic of Poland

CRITICAL THINKING OF UNIVERSITY STUDENTS IN THE ERA OF ARTIFICIAL INTELLIGENCE

Abstract. The rapid expansion of artificial intelligence (AI) tools in higher education has reopened fundamental questions about the trajectory of students' critical thinking (CT) development. This article reports the results of a longitudinal cohort diagnostic study conducted at three higher education institutions in Astana, Republic of Kazakhstan L.N. Gumilyov Eurasian National University, Kazakh National University of Sports, and Saken Seifullin Kazakh Agro Technical Research University across four measurement points: 2020, 2022, 2024, and 2026, with a total sample of 334 students. CT was assessed using the Starkey Critical Thinking Test, adapted for Russian-language audiences, which measures analytical reasoning, argument evaluation, fallacy identification, evidence assessment, and problem-solving capacity. Inferential analyses employed the Mann–Whitney U test, Cohen's d effect sizes, and the Kruskal–Wallis omnibus test. Three research hypotheses were formulated and tested: that CT levels would demonstrate a statistically significant positive trajectory over the study period; that the period of active AI adoption (2023–2024) would coincide with a peak in CT performance; and that subsequent normalisation of AI use would be associated with a partial decline. The results confirm statistically significant improvement from 2020 ($M = 14.07$, $SD = 3.71$) to 2024 ($M = 18.58$, $SD = 4.44$), with a large effect size ($d = 1.10$, $p < 0.001$), and a statistically non-significant partial decline in 2026 ($M = 17.03$, $p = 0.059$). The overall Kruskal–Wallis test confirms significant between-cohort differences ($H = 43.13$, $p < 0.001$). The findings are discussed in relation to active pedagogical strategies, media literacy, and the dual role of AI as both a stimulus and a potential inhibitor of critical reasoning. Implications for competency-based curriculum reform within the Bologna Process framework are considered.

Keywords: critical thinking; higher education; longitudinal study; Starkey test; artificial intelligence; media literacy; active learning; Bologna Process; Kazakhstan; competency-based education.

Introduction

The question of whether university education genuinely develops students' capacity for critical, independent reasoning has acquired new urgency in the context of digitalisation and the widespread adoption of generative artificial intelligence tools. When AI systems can produce coherent, persuasive text on virtually any academic topic within seconds, the value added by higher education can no longer rest on information transmission alone. It must reside, above all, in the cultivation of the capacity to evaluate, question, and reason that is, in critical thinking (CT).

Kazakhstani higher education has been navigating this challenge within the broader context of Bologna Process-aligned reform. The State Programme for the Development of Education and Science of the Republic of Kazakhstan (2020–2025) explicitly designates critical thinking as one of the core broad-spectrum skills that university graduates must demonstrate, alongside research competence, ICT literacy, communicative flexibility, and problem-solving capacity (Government of the Republic of Kazakhstan, 2019). This programmatic commitment was deepened and extended by the Concept for the Development of Pre-School, Secondary, Technical and Vocational Education of the Republic of Kazakhstan for 2023–2029. The Concept reorients national educational policy from remedying systemic deficits toward placing the learner and their

future competencies at the centre, explicitly listing critical thinking, global competencies, and emotional intelligence as the defining outcomes of modernised curricula (Government of the Republic of Kazakhstan, 2023). In its revised edition, the Concept further stipulates the integration of artificial intelligence elements into state educational standards, with critical thinking and information literacy identified as the competencies to be cultivated through AI-enhanced learning environments. Yet the empirical evidence on whether CT is actually developing among students, and how that trajectory has been affected by the COVID-19 pandemic, post-pandemic pedagogical restructuring, and the emergence of generative AI, remains scarce particularly from Central Asian institutional contexts.

The present study directly addresses this gap. Drawing on a series of diagnostic measurements conducted at three higher education institutions in Astana: L.N. Gumilyov Eurasian National University (ENU), Kazakh National University of Sports (KNUS), and Saken Seifullin Kazakh Agro Technical Research University (KazATU), between 2020 and 2026, we report the first longitudinal cohort analysis of CT development spanning the pre-pandemic, post-pandemic, early AI, and normalised AI periods in this regional context. The study builds on the authors' earlier work on CT diagnostics and active learning strategies in Kazakhstani vocational and higher education (Bekbayeva et al., 2020; Bekbayeva et al., 2022; Bekbayeva & Seidina, 2025), extending it with new data and a comparative longitudinal framework.

Three research hypotheses structure the inquiry. H1: CT levels among students will demonstrate a statistically significant positive trajectory over the 2020–2026 period, reflecting cumulative effects of competency-oriented pedagogical reform. H2: The period coinciding with the initial adoption of generative AI tools (2023–2024) will mark a peak in CT performance, reflecting the stimulating effect of novel technological engagement on analytical reasoning. H3: Subsequent normalisation of AI use (2025–2026) will be associated with a partial, though not necessarily significant, decline in CT scores, reflecting a shift from active to passive AI engagement patterns.

Literature Review

Critical Thinking: Conceptual Foundations

Critical thinking in higher education has been conceptualised along three main dimensions in the contemporary literature. The cognitive dimension encompasses discrete skills including analysis, inference, evaluation, and argumentation that allow students to assess the quality of information and construct logically warranted conclusions (Yessekeshova et al., 2020; Bezanilla et al., 2021). The metacognitive dimension involves the conscious regulation of one's own reasoning processes: monitoring for bias, recognising gaps in understanding, and adjusting cognitive strategies in response to feedback (Rivas et al., 2022). The dispositional dimension refers to stable intellectual orientations as curiosity, open-mindedness, and epistemic humility that determine whether a student who possesses CT skills actually applies them spontaneously in novel contexts (Álvarez-Huerta et al., 2023).

This three-part framework is particularly significant for the Kazakhstani context, where educational reform has traditionally emphasised knowledge acquisition over the cultivation of evaluative and reflective capacities. The competency-based approach introduced through Bologna Process alignment requires that graduates not merely know content but demonstrate the ability to apply knowledge critically, handle uncertainty, and make informed judgements, all of which presuppose well-developed CT in all three dimensions (Temirgazina et al., 2024).

Factors Shaping Critical Thinking Development

The empirical literature identifies pedagogical, psychological, sociocultural, and technological factors as the principal determinants of CT outcomes in university students. Among pedagogical factors, the most robustly evidenced approaches include problem-based learning (PBL), multi-level task design, inquiry-based learning, and case-based discussion (Yu & Zin, 2023; Bekbayeva et al., 2022). Multi-level task structures in particular where the cognitive demand of tasks is calibrated to challenge students progressively across analysis, synthesis, and evaluation have been shown to produce measurable CT gains among university students (Bekbayeva et al., 2021).

Media resources constitute an additional pedagogical lever whose potential for CT development has received growing attention. The use of audio-visual materials, digital texts, and media content as objects of critical analysis requiring students to evaluate source credibility, identify persuasive strategies, and assess

evidence quality engages multiple CT skill dimensions simultaneously. Bekbayeva and Seidina (2025), in a study of 218 students at a Kazakhstani university using a multi-level media task design, demonstrated statistically significant CT improvements (Cramer–Welch test), confirming the effectiveness of this approach for developing analytical, synthesis, and evaluation competencies in Russian-language academic contexts.

Among psychological factors, metacognition has emerged as the most consistent predictor of CT growth. Rivas, Saiz, and Ossa (2022) demonstrated through controlled experimentation that metacognitive self-regulation training produces substantially larger CT improvements than skills-focused instruction alone, establishing metacognition as a causal mechanism rather than merely a correlate of CT performance. Guamanga, Saiz, Rivas, and Morales Bueno (2025) further showed, using structural equation modelling, that CT predicts metacognition, which in turn predicts psychological well-being and empathy positioning CT as a core personal quality with implications beyond academic performance.

Technological factors present a more complex and contested picture. The systematic review by Melisa et al. (2025), covering AI's effects on CT in higher education, identified a paradoxical dynamic: students who engage critically with AI-generated content as questioning, fact-checking, and evaluating it tend to develop stronger CT skills, while those who use AI primarily to replace independent analytical effort show diminished reasoning capacity over time. This paradox is further elaborated in the bibliometric analysis of Salido et al. (2025), which mapped five thematic clusters in the AI-and-CT literature and identified pedagogical innovation and AI literacy as the two areas of most rapid growth since the widespread adoption of ChatGPT in 2023.

CT Diagnostics: Instruments and Approaches

The assessment of CT in empirical research relies on a combination of standardised psychometric instruments and performance-based tasks. Among internationally validated measures, the Watson-Glaser Critical Thinking Appraisal, the California Critical Thinking Skills Test, and the PENCRIASAL (developed by Rivas and Saiz at the University of Salamanca) are most widely used in Scopus- and WoS-indexed research (Araújo et al., 2024; Wang & Abdullah, 2024). A critical limitation of these instruments in the Central Asian context is the absence of validated Russian-language adaptations with established norms for Kazakhstani university populations.

The Starkey Critical Thinking Test, adapted for Russian-language audiences and employed in the present study, addresses this gap. It is a 27-item multiple-choice instrument assessing applied reasoning across five competency areas: argument recognition and evaluation, fallacy identification, evidence assessment, source credibility evaluation, and solution analysis. Its psychometric properties in the Kazakhstani context have been established in earlier work by the research team (Bekbayeva & Seidina, 2025; Bekbayeva et al., 2020), providing the methodological continuity necessary for longitudinal comparison. Score interpretation follows three levels: Low (0–9 points), Medium (10–18 points), and High (19–27 points).

Research Gap and Contribution

Despite significant international activity, CT research from Central Asian higher education contexts remains substantially underrepresented in high-impact indexed publications (Andreucci-Annunziata et al., 2023). The present study addresses this gap in three ways: by providing longitudinal diagnostic data spanning six years and four distinct technological-pedagogical phases; by employing a locally validated instrument with established comparability across measurement points; and by situating empirical findings within the theoretical and policy frameworks most relevant to Kazakhstani higher education as the Bologna Process competency agenda and the national education development programme.

Methodology

Research Design and Rationale

The study employs a longitudinal cohort design with four independent cross-sectional measurement points: October 2020, May 2022, December 2024, and March 2026. This design comparing distinct student cohorts at successive timepoints rather than tracking the same individuals allows examination of how CT levels among incoming university students have shifted across successive academic generations, each encountering a different technological and pedagogical environment. While individual-level longitudinal tracking would permit stronger causal claims, the cohort design is appropriate for the research objectives and consistent with practice in educational trend research (Wang & Abdullah, 2024).

The four measurement points correspond to four analytically distinct contextual phases: the pre-pandemic

baseline (2020); the early post-pandemic period, marked by rapid transition to blended learning (2022); the first year of widespread generative AI availability in Kazakhstan (2024); and the period of normalised AI integration (2026). This temporal structure makes the dataset uniquely suited to examining the hypothesised relationships between CT development and technological context.

Participants

The total sample comprised 334 students across four cohorts: 2020 ($n = 91$; female = 54, male = 37), 2022 ($n = 48$), 2024 ($n = 59$; female = 32, male = 27), and 2026 ($n = 136$; female = 116, male = 20). Participants were recruited from three higher education institutions in Astana, Republic of Kazakhstan: L.N. Gumilyov Eurasian National University, the Kazakh National University of Sports (KNUS), and Saken Seifullin Kazakh Agro Technical Research University (KazATU). All three institutions have been implementing Bologna Process-aligned competency-based curricula, making the multi-institutional sample representative of the broader Kazakhstani higher education reform context. Students were enrolled in first and second years of undergraduate programmes spanning agricultural, natural science, and pedagogical fields. The age range was predominantly 18–21 years. Participation was entirely voluntary; students were informed of the study objectives and guaranteed full confidentiality of their individual results. The notable gender imbalance in the 2026 cohort (85.3% female) reflects programme-specific enrolment patterns at the participating institutions in that academic year and is acknowledged as a potential confounding variable in the interpretation of results.

Instrument

The Starkey Critical Thinking Test is a 27-item multiple-choice instrument assessing applied reasoning in realistic everyday and academic scenarios. Each item presents a situation and requires the student to identify the most logically defensible response from four alternatives. Items target five competency areas: argument recognition and evaluation (items 1–6); fallacy identification, including ad hominem, post hoc reasoning, false dichotomy, and circular reasoning (items 7–13); evidence quality assessment (items 14–18); source credibility evaluation (items 19–22); and problem identification and solution analysis (items 23–27). Correct responses are scored 1 point each; total scores range from 0 to 27. The test was administered digitally through a Google Forms interface, with individual responses and total scores recorded automatically. The Russian-language adaptation used in this study has been applied consistently across all four measurement points, ensuring instrument comparability (Bekbayeva & Seidina, 2025).

Statistical Analysis

Data were processed using Python 3.11 (scipy.stats library). For each cohort, descriptive statistics were computed including mean (M), standard deviation (SD), standard error (SE), median, minimum, and maximum. Distribution normality was assessed using the Shapiro–Wilk test. Since the 2024 ($W = 0.944$, $p = 0.009$) and 2026 ($W = 0.973$, $p = 0.008$) cohorts violated normality assumptions, all inferential comparisons used non-parametric methods. Pairwise between-cohort comparisons employed the Mann–Whitney U test (two-tailed, $\alpha = 0.05$). Effect sizes were computed as Cohen's d using pooled standard deviations, with conventional thresholds of small (0.2–0.5), medium (0.5–0.8), and large (> 0.8) following Cohen (1988). Between-group differences across all four cohorts were assessed using the Kruskal–Wallis H test.

Results And Discussion

Descriptive Statistics and Level Distribution

Table 1 presents the combined descriptive statistics and CT level distributions for all four cohorts. The data reveal a pronounced non-linear trajectory: an upward trend from 2020 to 2024, a peak in 2024, and a partial reversal in 2026. The mean score rose from 14.07 in 2020 to 18.58 in 2024 — an increase of 4.51 points, corresponding to a shift from the middle of the Medium level to the threshold of the High level. In 2026, the mean declined to 17.03, remaining above the 2022 value but below the 2024 peak.

Table 1*Descriptive statistics and CT level distribution by cohort (Starkey Critical Thinking Test, 0–27 points)*

Year	N	M	SD	Min	Max	Mdn	SE	Low, %	Med, %	High, %
2020	91	14.07	3.71	6	22	14.0	0.39	12.1	74.7	13.2
2022	48	15.42	3.61	7	22	16.0	0.52	4.2	75.0	20.8
2024	59	18.58	4.44	7	25	19.0	0.58	1.7	39.0	59.3
2026	136	17.03	5.28	2	27	17.0	0.45	10.3	47.8	41.9

The level distribution data amplify this picture. The proportion of students achieving the High CT level rose dramatically from 13.2% in 2020 to 59.3% in 2024 — a more than fourfold increase over four years. Simultaneously, the Low level proportion fell from 12.1% to 1.7%. In 2026, the High level proportion declined to 41.9% while the Low level partially recovered to 10.3%, suggesting a redistribution of students from the High level toward the Medium and Low categories. The standard deviation also increased from 3.71 (2020) to 5.28 (2026), indicating growing heterogeneity in CT outcomes — a pattern consistent with diverging student engagement patterns under conditions of widespread AI availability.

Hypothesis Testing: Inferential Results

The omnibus Kruskal–Wallis test confirmed statistically significant overall differences across all four cohorts ($H = 43.13$, $p < 0.001$), warranting pairwise comparisons. Table 2 presents all six pairwise Mann–Whitney U comparisons with associated statistics.

Table 2*Pairwise Mann–Whitney U test results with effect sizes ($\alpha = 0.05$)*

Comparison	U	p	Sig.	Cohen's d	Effect	ΔM	Direction
2020 vs 2022	1730	0.044	*	0.37	Small	+1.35	↑
2020 vs 2024	1176	<0.001	***	1.10	Large	+4.51	↑
2020 vs 2026	3837	<0.001	***	0.65	Medium	+2.96	↑
2022 vs 2024	802	<0.001	***	0.78	Medium	+3.16	↑
2022 vs 2026	2485	0.014	*	0.36	Small	+1.61	↑
2024 vs 2026	4694	0.059	n.s.	0.32	Small	–1.55	↓

Note. Sig. = significance level: *** $p < 0.001$; * $p < 0.05$; n.s. = not significant. ΔM = difference in group means (positive = later group higher).

Hypothesis H1: that CT levels would demonstrate a statistically significant positive trajectory, is supported. All comparisons involving the 2020 baseline cohort reached statistical significance. The comparison between 2020 and 2024 produced the largest observed effect in the dataset ($d = 1.10$), qualifying as large by conventional benchmarks and representing the most robust evidence of CT improvement across the study period. Importantly, even the comparison between 2020 and 2026 spanning the full six-year range remains highly significant ($p < 0.001$, $d = 0.65$, medium effect), confirming that CT gains were not entirely reversed by 2026.

Hypothesis H2: that the 2024 cohort would represent a peak in CT performance, is also supported. The 2024 cohort recorded the highest mean score (18.58) and the largest proportion of High-level students (59.3%) across all measurement points. The comparison between 2022 and 2024 yielded a medium effect ($d = 0.78$, $p < 0.001$), confirming that 2024 performance was substantially superior to the previous measurement point, not merely to the 2020 baseline.

Hypothesis H3: that the 2026 cohort would show a partial decline, receives qualified support. The comparison between 2024 and 2026 shows a mean difference of –1.55 points in the expected direction, but this difference does not reach statistical significance ($p = 0.059$). Strictly interpreted, H3 cannot be confirmed at the 0.05 level. However, the p-value of 0.059 represents a borderline result, and the observed decline is substantively meaningful, particularly when combined with the structural changes in level distribution (High

level falling from 59.3% to 41.9%; Low level recovering from 1.7% to 10.3%). This pattern warrants continued monitoring and does not permit dismissal of the underlying dynamic.

Interpreting the Longitudinal Trajectory

The overall pattern presents significant rise to 2024, borderline non-significant partial reversal in 2026 and admits several interpretations that are not mutually exclusive and may operate simultaneously.

The first interpretation centres on pedagogical restructuring. The period 2020–2024 was characterised by rapid institutional adaptation to online and blended learning, driven by pandemic necessity. This transition compelled faculty to redesign courses around interactive, problem-oriented formats, precisely the active learning approaches that the literature consistently identifies as most effective for CT development (Yu & Zin, 2023; Bekbayeva et al., 2022). The introduction of media-based multi-level tasks and structured critical analysis activities approaches empirically validated in the Kazakhstani context (Bekbayeva & Seidina, 2025) likely contributed directly to the observed improvements. The partial decline in 2026 may reflect a reversion toward conventional instructional formats as the emergency stimulus of the pandemic receded.

The second interpretation, which we term the AI novelty hypothesis, draws on the paradox identified in the systematic review literature (Melisa et al., 2025; Salido et al., 2025). When ChatGPT and similar tools became widely accessible to Kazakhstani students in early 2023, the novelty of the technology prompted active engagement: students interrogated AI outputs, compared them with other sources, questioned their accuracy, and discussed their limitations. This metacognitive engagement with novel AI content may have inadvertently stimulated the same analytical processes that CT instruction aims to cultivate. By 2026, however, two years of normalised AI use may have shifted many students toward more passive consumption patterns — accepting AI outputs without critical evaluation, delegating analytical tasks to AI systems, and thereby reducing the frequency of independent reasoning exercises.

The third interpretation addresses sampling variability. The 2026 cohort was substantially larger ($n = 136$) and exhibited a markedly different gender composition (85.3% female) compared with earlier cohorts. If gender is associated with differences in CT performance in the local context, a question that was not the focus of this study and cannot be resolved from available data, this compositional shift could partially account for the lower mean score. The wider standard deviation in 2026 (5.28 vs. 4.44 in 2024) is also consistent with a more heterogeneous intake, potentially reflecting broader programme-level enrolment patterns rather than a systematic CT decline.

The fourth interpretation posits stabilisation rather than decline: that the sustainable CT equilibrium under current institutional conditions lies in the range of 16–18 points, with the 2024 peak reflecting a transient confluence of favourable conditions: intensive pedagogical restructuring, technology novelty effect, and heightened student motivation, unlikely to be replicated without deliberate and sustained intervention. This interpretation is consistent with the non-significant p-value for the 2024–2026 comparison and suggests that the primary research task is not to explain a decline but to identify the conditions that enabled the 2024 peak and determine how to sustain them.

Across all four interpretations, a common implication emerges: CT development in higher education is neither automatic nor irreversible. It depends on active pedagogical investment, intentional engagement with analytical tasks, and the cultivation of metacognitive habits that enable students to monitor and improve their own reasoning. Psychological and pedagogical guidance approaches that explicitly address these dimensions including coaching, reflective practice, and deliberate AI-critical literacy training offer a promising pathway (Phembayev & Uaikhanova, 2024). The competency model of a Kazakhstani pedagogical university graduate, as defined in state regulatory documents, provides the normative framework within which such approaches should be embedded (Temirgazina et al., 2024).

Implications for Policy and Practice

The findings carry several concrete implications for Kazakhstani higher education institutions operating within the Bologna Process framework. First, the significant positive trajectory from 2020 to 2024 demonstrates that CT improvement is achievable at the institutional level within meaningful timeframes, provided that pedagogical conditions actively support analytical reasoning. This should encourage investment in active learning infrastructure, faculty development in CT pedagogy, and the systematic integration of multi-level and media-analytical task designs across disciplinary curricula.

Second, the borderline decline in 2026, combined with the structural changes in level distribution, signals

the need for proactive AI literacy integration into CT pedagogy. Rather than treating AI as a threat to student reasoning capacity, institutions should design learning environments in which AI outputs serve as objects of critical analysis as stimuli for evaluation, comparison, and argumentation, rather than as substitutes for independent thought. This pedagogical reframing of AI aligns directly with the CT competency goals of the State Programme for the Development of Education and Science.

Third, the growing heterogeneity of CT outcomes (reflected in widening standard deviations) suggests that average scores increasingly mask divergent student trajectories. Diagnostic monitoring should be accompanied by differentiated pedagogical support: targeted intervention for students at the Low level, challenge-based extension for those at the High level, and structured metacognitive scaffolding for the majority in the Medium range.

Conclusion

This longitudinal cohort study of critical thinking among university students in Kazakhstan, spanning four measurement points from 2020 to 2026 and involving 334 participants, yields findings of both theoretical and practical significance. The data confirm a substantial and statistically robust improvement in CT from the pre-pandemic baseline to the period of initial AI adoption, with the 2024 cohort recording the highest mean score (18.58 out of 27) and the largest proportion of High-level students (59.3%) across the entire observation period. A borderline-significant partial decline in 2026 raises important questions about the sustainability of CT gains under conditions of normalised AI use, though the 2026 cohort still significantly outperforms the 2020 baseline.

Three hypotheses were formulated and tested. H1, predicting a significant positive trajectory over the study period, is fully supported. H2, predicting that the 2024 cohort would represent a performance peak, is supported. H3, predicting a partial decline in 2026, receives qualified support: the decline is substantively meaningful and in the predicted direction, but does not reach conventional statistical significance at the 0.05 threshold.

The study contributes original empirical evidence from a Kazakhstani higher education context that has been underrepresented in international CT research. By employing a locally validated instrument (the Starkey Critical Thinking Test in its Russian-language adaptation) consistently across all measurement points, and by situating findings within both the theoretical literature on CT and the policy framework of Kazakhstani educational reform, the study offers a model for systematic longitudinal CT monitoring at the institutional level.

The limitations of the study are acknowledged: the cross-sectional cohort design precludes individual-level causal inference; the geographically concentrated multi-institutional sample limits transferability; the absence of AI usage data prevents direct testing of the AI paradox hypothesis; and the gender imbalance in the 2026 cohort introduces potential confounding. Future research should address these limitations through longitudinal individual tracking, multi-institutional sampling, integration of AI engagement pattern measurement, and gender-stratified analysis. The ongoing research programme of the present authors, building on this diagnostic foundation, will extend the analysis to include intervention studies and qualitative exploration of student reasoning processes.

Conflict Of Interest Statement

The authors declare no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Author Contributions

Pawel Bogacz: International Context and Comparative Perspective, Contribution of Information on International Students, Review and Editing. Zhanar Bekbayeva: Writing – Original Draft Preparation, Development of the Main Section of the Article. Moldir Seidina: Investigation, Data Collection and Research Conducted at Her University, Review and Editing. Saule Aldabergenova: Investigation, Data Collection and Research Conducted at Her University, Review and Editing. Botagoz Bermukhambetova: Article Formatting, Investigation in the Area of Digital Competence Development, Writing – Review and Editing.

Declaration of Use of Generative AI and AI-Assisted Technologies

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript.

References

- Álvarez-Huerta, P., Muela, A., & Larrea, I. (2023). Disposition towards critical thinking and student engagement in higher education. *Innovative Higher Education*, 48(2), 239–256. <https://doi.org/10.1007/s10755-022-09614-9>
- Andreucci-Annunziata, P., Riedemann, A., Cortés, S., Mellado, A., del Río, M. T., & Vega-Muñoz, A. (2023). Conceptualizations and instructional strategies on critical thinking in higher education: A systematic review of systematic reviews. *Frontiers in Education*, 8, 1141686. <https://doi.org/10.3389/educ.2023.1141686>
- Araújo, B., Gomes, S. F., & Ribeiro, L. (2024). Critical thinking pedagogical practices in medical education: A systematic review. *Frontiers in Medicine*, 11, 1358444. <https://doi.org/10.3389/fmed.2024.1358444>
- Bekbayeva, Z. S., & Seidina, M. Z. (2025). Using media resources in education to develop university students' critical thinking skills. *ILIM*, 43(1), 8–17. <https://doi.org/10.47751/skpu.1937.v43i1.1>
- Bekbayeva, Z. S., Solovyev, A. N., Chechetkina, N. V., Bolshakov, A. A., & Sosina, L. V. (2022). Problems of organizing critical thinking training for university students. *Journal of Higher Education Theory and Practice*, 22(10), 116–123. <https://doi.org/10.33423/jhetp.v22i10.5390>
- Bekbayeva, Z. S., Galiyev, T. T., Albytova, N., Zhazykbayeva, Z. M., & Mussatayeva, A. B. (2021). Fostering post-secondary vocational students' critical thinking through multi-level tasks in learning environments. *World Journal on Educational Technology: Current Issues*, 13(3), 397–406. <https://doi.org/10.18844/wjet.v13i3.5948>
- Bezanilla, M. J., Fernández-Nogueira, D., Poblete, M., & Galindo-Domínguez, H. (2021). Methodologies for teaching-learning critical thinking in higher education: The teacher's view. *Thinking Skills and Creativity*, 42, 100892. <https://doi.org/10.1016/j.tsc.2021.100892>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Government of the Republic of Kazakhstan. (2019). State Programme for the Development of Education and Science of the Republic of Kazakhstan for 2020–2025 [Государственная программа развития образования и науки Республики Казахстан на 2020–2025 годы]: Government Resolution No. 988. Expired upon completion of implementation cycle. <https://www.gov.kz/memleket/entities/edu/documents/details/14703?lang=ru>
- Government of the Republic of Kazakhstan. (2023). Concept for the Development of Pre-School, Secondary, Technical and Vocational Education of the Republic of Kazakhstan for 2023–2029 [Концепция развития дошкольного, среднего, технического и профессионального образования Республики Казахстан на 2023–2029 годы]: Government Resolution No. 249 of 28 March 2023 (as amended by Resolution No. 465 of 13 June 2024; currently in force). <https://adilet.zan.kz/rus/docs/P2300000249>
- Guamanga, M. H., Saiz, C., Rivas, S. F., & Morales Bueno, P. (2025). Critical thinking and metacognition: Pathways to empathy and psychological well-being. *Journal of Intelligence*, 13(3), 34. <https://doi.org/10.3390/jintelligence13030034>
- Jaramillo Gómez, D. L., Álvarez Maestre, A. J., Parada Trujillo, A. E., Pérez Fuentes, C. A., Bedoya Ortiz, D. H., & Sanabria Alarcón, R. K. (2025). Determining factors for the development of critical thinking in higher education. *Journal of Intelligence*, 13(6), 59. <https://doi.org/10.3390/jintelligence13060059>
- Melisa, R., Ashadi, A., Triastuti, A., Hidayati, S., Salido, A., Ero, P. E. L., Marlini, C., Zefrin, Z., & Al Fuad, Z. (2025). Critical thinking in the age of AI: A systematic review of AI's effects on higher education. *Educational Process: International Journal*, 14, e2025031. <https://doi.org/10.22521/edupij.2025.14.31>
- Phembayev, M. A., & Uaikhanova, M. (2024). Comparative analysis of foreign experience in implementing psychological and pedagogical guidance for soft skills development of university students. *Higher Education in Kazakhstan*, 47(3), 79–86. <https://doi.org/10.59787/2413-5488-2024-47-3-79-86>

- Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology*, 13, 913219. <https://doi.org/10.3389/fpsyg.2022.913219>
- Salido, A., Ero, P. E. L., Melisa, R., Marlini, C., Ashadi, A., Triastuti, A., Hidayati, S., Zefrin, Z., & Al Fuad, Z. (2025). Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. *Social Sciences & Humanities Open*, 12, 101924. <https://doi.org/10.1016/j.ssaho.2025.101924>
- Temirgazina, Z., Andryuchshenko, O., & Sirlibaev, M. (2024). Reflection of the competence model of a pedagogical university graduate in regulative documents. *Higher Education in Kazakhstan*, 46(2), 118–127. <https://doi.org/10.59787/2413-5488-2024-46-2-118-127>
- Wang, Q., & Abdullah, A. H. (2024). Enhancing students' critical thinking through mathematics in higher education: A systematic review. *SAGE Open*, 14(3). <https://doi.org/10.1177/21582440241275651>
- Yessekeshova, M. D., Ibrayeva, K. Zh., Kaltayeva, G. K., Albytova, N., & Bekbayeva, Z. S. (2020). On the formation of critical thinking of students of a higher educational institution. *Journal of Intellectual Disability – Diagnosis and Treatment*, 8(3), 388–395. <https://doi.org/10.6000/2292-2598.2020.08.03.15>
- Yu, L., & Zin, Z. M. (2023). The critical thinking-oriented adaptations of problem-based learning models: A systematic review. *Frontiers in Education*, 8, 1139987. <https://doi.org/10.3389/feduc.2023.1139987>

Information about authors

Zhanar Bekbayeva – PhD, Senior Lecturer, Department of Pedagogy, L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan; e-mail: bekbayeva@inbox.ru; ORCID: 0000-0002-6070-4195.

Moldir Seidina – PhD, Associate Professor, Department of Management and Innovation in Sports, Kazakh National University of Sports, Astana, Republic of Kazakhstan; e-mail: m_seidina@apems.edu.kz; ORCID: 0000-0002-9106-9147.

Saule Aldabergenova – PhD, Head of Educational Programs in Transport Engineering, Technology, Standardization, Certification, and Metrology, S. Seifullin Kazakh Agrotechnical Research University, Astana, Republic of Kazakhstan; e-mail: s.algabergenova@kazatu.edu.kz; ORCID: 0000-0001-6297-7632.

Botagoz Bermukhambetova – Deputy Director, Higher Education Development National Center of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Republic of Kazakhstan; e-mail: b.bermukhambetova@n-k.kz; ORCID: 0000-0003-1414-6484. (**corresponding author**)

Pawel Bogacz – DSc, PhD Eng., Professor at AGH University of Krakow, Department of Economics and Management in Industry, Faculty of Civil Engineering and Resource Management, Krakow, Republic of Poland; e-mail: bogacz@agh.edu.pl; ORCID: 0000-0002-0099-6376.

Received: 02.07.2026

Revised: 31.08.2026

Accepted: 21.09.2026

Published: 30.09.2026

Type: Research Article