ISSN: 1991-3494 (Print)
ISSN: 2518-1467 (Online)

SCIENTIFIC JOURNAL OF PEDAGOGY AND ECONOMICS

№2
2026

ISSN 2518-1467 (Online),
ISSN 1991-3494 (Print)

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

SCIENTIFIC JOURNAL OF PEDAGOGY AND ECONOMICS

PUBLISHED SINCE 1944

2 (420)

March – April 2026

ALMATY, 2026

Scientific Journal of Pedagogy and Economics
ISSN 2518-1467 (Online),
ISSN 1991-3494 (Print).
Owner: «Central Asian Academic Research Center» LLP (Almaty).
The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Communications of the Republic of Kazakhstan
№ KZ50VPY00121155, issued on 05.06.2025
Thematic focus: *«publication of the results of new achievements inthe field of fundamental sciences»*
Periodicity: 6 times a year.
http://www.bulletin-science.kz/index.php/en/

Scientific Journal of Pedagogy and Economics
ISSN 2518-1467 (Online),
ISSN 1991-3494 (Print).
Меншіктенуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы қ.).
Қазақстан Республикасының Ақпарат және коммуникациялар министрлігінің Ақпарат комитетінде 05.06.2025 ж. берілген **№ KZ50VPY00121155** мерзімдік басылым тіркеуіне қойылу туралы куәлік.
Тақырыптық бағыты: *«іргелі ғылым салалары бойынша жаңа жетістіктердің нәтижелерін жариялау»*
Мерзімділігі: жылына 6 рет.
http://www.bulletin-science.kz/index.php/en/

Scientific Journal of Pedagogy and Economics
ISSN 2518-1467 (Online),
ISSN 1991-3494 (Print).

Собственник: ТОО «Центрально-азиатский академический научный центр» (г. Алматы).

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и коммуникаций и Республики Казахстан

Nº KZ50VPY00121155 выданное 05.06.2025 г.

Тематическая направленность: *«публикация результатов новых достижений вобласти фундаментальных наук».*

Периодичность: 6 раз в год.

http://www.bulletin-science.kz/index.php/en/

CONTENTS

PEDAGOGY

ECONOMICS

МАЗМҰНЫ

ПЕДАГОГИКА

ЭКОНОМИКА

СОДЕРЖАНИЕ

ПЕДАГОГИКА

ЭКОНОМИКА

SCIENTIFIC JOURNAL OF PEDAGOGY AND ECONOMICS
ISSN 1991-3494
Volume 2.
Number 420 (2026), 190-207

https://doi.org/10.32014/2026.2518-1467.1154

UDC: 372.854
IRSTI: 14.25.09

© Kabzhalelov K.R.*, Korganbayeva Zh.K., Nurakhmetova A.R., 2026.
Abai Kazakh National Pedagogical University, Almaty, Kazakhstan.
E-mail: kamil_kabzhalelov@outlook.com

DEVELOPING CRITICAL THINKING THROUGH MODERN CHATBOTS IN CHEMISTRY EDUCATION

Kabzhalelov Kamil — 1-st-year PhD Student, Faculty of Natural Sciences and Geography, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan,
E-mail: kamil_kabzhalelov@outlook.com, ORCID: https://orcid.org/0009-0008-6030-052X;
Korganbayeva Zhanar — Candidate of Chemical Sciences, Associate Professor, Faculty of Geography and Natural Sciences, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan,
E-mail: korganbaeva.zhan@mail.ru, ORCID: https://orcid.org/0000-0001-8046-9586;
Nurakhmetova Aigul — Candidate of Pedagogical Sciences, Associate Professor, Faculty of Natural Sciences and Geography, Abai Kazakh National Pedagogical University, Kazakhstan, Almaty,
E-mail:Nurahmetovaa@bk.ru, ORCID: https://orcid.org/0000-0001-8486-2158.

Abstract. This article examines the pedagogical potential of modern AI-powered chatbots, specifically ChatGPT, for developing students' critical thinking in chemistry classes. The relevance of the study stems from the rapid digitalization of education and the growing volume of automatically generated information, which requires students to develop skills for conscious analysis, interpretation, and critical evaluation of data obtained from AI sources. The aim of the study was to determine the effectiveness of using educational texts created using ChatGPT and containing intentional logical and factual errors as a means of developing critical thinking in chemistry lessons. The study was conducted as a quasi-experiment with a pre-test/post-test design in grades 8 and 9 of a comprehensive school in Almaty (N = 58) during the third quarter. The experimental groups systematically used texts containing intentional errors, generated using ChatGPT and pre-edited by the teacher, while the control group used the problem-based presentation method. The level of critical thinking was assessed before and after the intervention using a modified L. Lutsenko method; statistical analysis included the Shapiro-Wilk test, Student's t-test, Wilcoxon and Mann-Whitney tests, and Cohen's d effect size calculation. The results showed a statistically significant increase in critical thinking scores in the experimental groups (d = 1.88 and d = 0.75) with minimal changes in the control group (d = 0.23). The data obtained indicate that the analysis of texts with intentional errors contributes to increased cognitive activity and the development of critical information assessment skills. The theoretical significance lies in the development of a reproducible algorithm for generating educational texts with intentional errors using ChatGPT. The practical significance lies in the possibility of using the developed materials as an additional educational tool for testing knowledge and stimulating analytical thinking in students in chemistry lessons.

Keywords: critical thinking, artificial intelligence, ChatGPT, chemistry education, prompt engineering

For citations: Kabzhalelov K.R., Korganbayeva Zh.K., Nurakhmetova A.R.. Developing critical thinking through modern chatbots in chemistry education. Scientific journal of pedagogy and economics, 2026. — No.2. – P. 190-207. DOI: https://doi.org/10.32014/2026.2518-1467.1154

© Кабжалелов К.Р.*, Қорғанбаева Ж.Қ., Нурахметова А.Р., 2026.
Абай атындағы Қазақ ұлттық педагогикалық университеті,
Алматы, Қазақстан.
E-mail: kamil_kabzhalelov@outlook.com

ХИМИЯНЫ ОҚЫТУ БАРЫСЫНДА ЗАМАНАУИ ЧАТ-БОТТАР АРҚЫЛЫ СЫНИ ОЙЛАУДЫ ДАМЫТУ

Кабжалелов Камиль — 1-курс докторанты, Жаратылыстану және география факультеті, Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан,
E-mail: kamil_kabzhalelov@outlook.com, ORCID: https://orcid.org/0009-0008-6030-052X;
Қорғанбаева Жанар — химия ғылымдарының кандидаты, қауымдастырылған профессор, География және жаратылыстану факультеті, Абай атындағы ҚАЗҰПУ, Алматы, Қазақстан,
E-mail: korganbaeva.zhan@mail.ru, ORCID: https://orcid.org/0000-0001-8046-9586;
Нурахметова Айгуль — педагогика ғылымдарының кандидаты, доцент, Жаратылыстану және география факультеті, Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан,
E-mail: Nurahmetovaa@bk.ru, ORCID: https://orcid.org/0000-0001-8486-2158.

Аннотация: Мақалада химияны оқыту барысында оқушылардың сыни ойлауын дамыту үшін жасанды интеллект негізіндегі заманауи чат-боттардың, атап айтқанда ChatGPT-дің педагогикалық мүмкіндіктері қарастырылады. Білім беруді белсенді цифрландыру және автоматты түрде жасалатын ақпарат көлемінің артуы жағдайында мектеп оқушыларында ЖИ-көздерінен алынатын деректерді саналы түрде талдау, интерпретациялау және сыни тұрғыдан бағалау дағдыларын қалыптастыру қажеттілігі зерттеудің өзектілігін айқындайды. Жұмыстың мақсаты — химия сабақтарында сыни ойлауды дамыту құралы ретінде ChatGPT көмегімен жасалған және құрамында әдейі жіберілген логикалық және фактілік қателері бар оқу мәтіндерін пайдаланудың тиімділігін анықтау болып табылады. Зерттеу үшінші тоқсан бойы Алматы қаласындағы жалпы білім беретін мектептің 8 және 9-сынып оқушыларының (N = 58) қатысуымен pre-test/post-test дизайны бойынша квазиэксперимент түрінде жүргізілді. Эксперименттік топтарда ChatGPT арқылы жасалып, мұғалім тарапынан алдын ала өңделген, әдейі қателері бар мәтіндер жүйелі түрде қолданылса, бақылау тобында проблемалық баяндау әдісі пайдаланылды. Оқушылардың сыни ойлау деңгейі Л. Луценконың модификацияланған әдістемесі бойынша интервенцияға дейін және одан кейін диагностикаланып, алынған деректер Шапиро-Уилк, Стьюденттің t-тесті, Уилкоксон және Манн-Уитни критерийлері, сондай-ақ Cohen's d эффект шамасы арқылы статистикалық өңдеуден өтті. Зерттеу нәтижелері бақылау тобындағы мардымсыз өзгерістермен (d=0,23) салыстырғанда, эксперименттік топтарда сыни ойлау көрсеткіштерінің статистикалық маңызды өсуін (d = 1,88 және d = 0,75) көрсетті. Алынған мәліметтер әдейі қателері бар мәтіндерді талдау оқушылардың танымдық белсенділігін арттыруға және ақпаратты сыни бағалау дағдыларын дамытуға ықпал ететінін дәлелдейді. Жұмыстың теориялық маңыздылығы ChatGPT көмегімен қатесі бар оқу мәтіндерін жасаудың қайта жаңғыртылатын алгоритмін әзірлеуде болса, практикалық құндылығы бұл материалдарды химия сабақтарында білімді тексеру мен талдамалық ойлауды ынталандырудың қосымша құралы ретінде пайдалану мүмкіндігімен айқындалады.

Түйін сөздер: сыни ойлау, жасанды интеллект, ChatGPT, химияны оқыту, сұраныс инженериясы

© Кабжалелов К.Р. *, Корганбаева Ж.К., Нурахметова А.Р., 2026.
Казахский Национальный педагогический университет имени Абая,
Алматы, Казахстан.
E-mail: kamil_kabzhalelov@outlook.com

РАЗВИТИЕ КРИТИЧЕСКОГО МЫШЛЕНИЯ С ПОМОЩЬЮ СОВРЕМЕННЫХ ЧАТ-БОТОВ ПРИ ОБУЧЕНИИ ХИМИИ

Кабжалелов Камиль — докторант 1 курса, Факультет естествознания и географии, Казахский национальный педагогический университет имени Абая, Алматы, Казахстан,
E-mail: kamil_kabzhalelov@outlook.com, ORCID: https://orcid.org/0009-0008-6030-052X;
Корганбаева Жанар — кандидат химических наук, ассоц. профессор, Факультет естествознания и географии, Казахский национальный педагогический университет имени Абая, Алматы, Казахстан,
E-mail: korganbaeva.zhan@mail.ru, ORCID: https://orcid.org/0000-0001-8046-9586;
Нурахметова Айгуль — кандидат педагогических наук, доцент, Факультет естествознания и географии, Казахский национальный педагогический университет имени Абая, Алматы, Казахстан,
E-mail: Nurahmetovaa@bk.ru, ORCID: https://orcid.org/0000-0001-8486-2158.

Аннотация: В статье рассматриваются педагогические возможности современных чат-ботов на основе искусственного интеллекта, в частности

ChatGPT, для развития критического мышления учащихся при обучении химии. Актуальность исследования обусловлена активной цифровизацией образования и ростом объёма автоматически генерируемого контента, что требует формирования у школьников навыков осознанного анализа, интерпретации и критической оценки информации, получаемой из источников, основанных на искусственном интеллекте. Целью исследования является определение эффективности использования учебных текстов, созданных с помощью ChatGPT и содержащих преднамеренные логические и фактические ошибки, в качестве средства развития критического мышления на уроках химии. Исследование проводилось в форме квазиэксперимента с дизайном pre-test/post-test в 8-х и 9-х классах общеобразовательной школы г. Алматы (N = 58) в течение третьей учебной четверти. В экспериментальных группах систематически применялись тексты с преднамеренными ошибками, сгенерированные с помощью ChatGPT и предварительно отредактированные учителем, тогда как в контрольной группе использовался метод проблемного изложения. Уровень критического мышления оценивался до и после педагогического вмешательства с использованием модифицированной методики Л. Луценко. Статистическая обработка данных включала критерий Шапиро–Уилка, t-тест Стьюдента, критерии Уилкоксона и Манна–Уитни, а также расчёт величины эффекта (Cohen's d). Результаты исследования показали статистически значимый рост показателей критического мышления в экспериментальных группах (d = 1,88 и d = 0,75) при минимальных изменениях в контрольной группе (d = 0,23). Полученные данные свидетельствуют о том, что анализ текстов с преднамеренными ошибками способствует повышению познавательной активности и развитию навыков критической оценки информации. Теоретическая значимость состоит в разработке воспроизводимого алгоритма генерации учебных текстов с преднамеренными ошибками с использованием ChatGPT. Практическая значимость определяется возможностью использования разработанных материалов в качестве дополнительного дидактического инструмента для проверки знаний и стимулирования аналитического мышления учащихся на уроках химии.

Ключевые слова: критическое мышление, искусственный интеллект, ChatGPT, обучение химии, инженерия запросов

Introduction. In today's world, developing critical thinking skills is a pressing educational challenge. To realize students' potential, it's crucial to not only address their academic knowledge but also recognize the importance of critical thinking and other vital skills (Townley, 2018) Many higher education institutions have included this skill in their educational objectives and introduced critical thinking courses as optional subjects (Zahavi and Friedman, 2020; Cruz et al., 2021; Ahern et al., 2012). Despite the enormous importance of this skill, there is still no single, precise definition that all scholars and educators agree upon (Van Damme et al., 2023).

A unified standard for assessing critical thinking has also not been developed, although numerous methods and tests are used, a review of which is presented in (Butler, 2024).

According to (Raj et al., 2022), critical thinking (CT) is the ability to think clearly and rationally about what to do and what to believe. Another definition focuses on the purposefulness and self-regulation of judgment, which leads to the interpretation, analysis, and evaluation of information. CT itself refers to higher-order thinking (Liu, Pásztor, 2022). It is generally accepted that critical thinking is a complex psychological construct consisting of skills and dispositions (attitudes). These two attributes of CT are interrelated; thanks to attitudes, an individual strives for and understands the importance of a critical approach to assessing information and solving problems. Skills, in turn, represent the direct tools for implementing critical thinking.

Literary review. The importance of CT in education lies in its ability to actively engage students in the learning process, foster a deep understanding of complex topics, and develop independence in both judgment and the organization of their own learning activities (Rivas et al., 2022). This skill enhances problem-solving and the ability to evaluate the reliability of information, which is especially important in a constantly changing world (Thornhill-Miller et al, 2023).

Globally, there are numerous examples of various approaches to developing critical thinking in teaching various disciplines. For example, in (Zulyusri et al., 2023), the authors propose problem-based learning as the primary method for activating student learning. According to their hypothesis, this approach not only promotes the development of problem-solving skills in everyday life but also simultaneously improves critical information processing skills. Problem-based learning is implemented primarily through case studies. The syntax of the project-based learning method is able to assist students in optimizing their creative and critical mindsets by starting learning with essential questions, working together to plan, developing project completion schedules, timelines, and deadlines (Zulyusri et al., 2023).

In a review article (Andreucci-Annunziata et al., 2023), the authors cite the most common techniques aimed at developing critical thinking in higher education institutions. These include debates, brainstorming, reflective writing, journaling, inductive and deductive reasoning training, lecture discussions, role-playing games, reflection methods, and peer assessment. Various brainstorming techniques help overcome cognitive and social biases (Al-Samarraie and Hurmuzan, 2018). Reflective writing contributes to the improvement of argumentation skills, self-assessment, and critical thinking (Sani et al, 2018).

With the development of artificial intelligence technologies, chatbots, and other tools based on neural networks and deep learning mechanisms, the ability to think critically is becoming especially important. AI-generated information requires careful and comprehensive human verification. The large-scale implementation of neural networks is also having a significant impact on education, transforming both student learning processes and teaching methods. Popular chatbots such as ChatGPT, Gemini, Claude, and others are already being actively used by school and university students, as well as teachers in their professional activities. According to M.K. Yelshina and co-authors (Elshina et al., 2025), AI is capable of creating individual educational trajectories, integrating gamification elements, acting as a virtual assistant and tutor, and providing feedback. Thus, AI tools have significant potential for the development of self-education.

It should be emphasized that, despite the significant potential of AI as a teaching tool, this innovative technology is associated with certain risks that can have a negative impact on the formation and development of critical thinking. A systematic literature review (Zhai et al., 2024) devoted to the study of the impact of modern chatbots on students' cognitive abilities identified several key patterns. The main and most frequently mentioned problem is that excessive reliance on dialogue with AI can negatively impact students' analytical skills and their ability to make decisions in difficult situations. Users begin to unquestioningly accept recommendations generated by AI. Regular use of chatbots can lead to a weakening of the ability to retain information. Dependence on these technologies can contribute to plagiarism and data falsification, jeopardizing academic integrity in both schools and universities (Kim et al., 2023). The problem of the generation of non-existent information, which is sometimes difficult to distinguish from the truthful, remains relevant (Hatem et al., 2023). Therefore, many leading companies have implemented a feature in their chatbots that requires mandatory references to sources used by neural networks.

Given the trends and risks outlined above, the development and implementation of pedagogical approaches aimed at fostering and developing critical thinking in students is particularly relevant. This skill is considered a key element when working with large amounts of information, including that generated by AI systems. In this regard, a pressing practical task is to find methods and techniques capable of integrating AI tools into the educational process not only as a source of information but also as a means of developing analytical skills. The purpose of this study is to develop and use assignments, namely, educational texts created using ChatGPT with intentional logical and substantive errors, to foster critical thinking in schoolchildren when teaching chemistry.

Research Questions of the Study:

- What is the initial level of students' critical thinking before the intervention?
- Does systematic work with AI-generated error-infused texts lead to significant improvements in critical thinking skills?
- Are there differences in outcomes between students exposed to the intervention and those taught using traditional problem-based instruction?
- What is the magnitude of the intervention's effect on students' critical thinking development?

The study is based on the hypothesis that educational texts created using ChatGPT will lead to a statistically significant increase in the level of critical thinking in 8th- and 9th-grade students in the experimental group compared to both their baseline level (according to Lutsenko's method) and the control group.

Materials and Methods. The study was conducted as a quasi-experiment during the third term of the 2024 - 2025 academic year (from January 21 to March 8) at Secondary School No. 69 in Almaty. The sample included students from three classes: two ninth-grade classes (9B - 19 students, 9V - 15 students) and one eighth-grade class (8B - 24 students). One ninth-grade class served as the control group, while the other was assigned as the experimental group. The eighth-grade class was used to examine the stability and transferability of the effect.

The research employed the following methods: analysis of scientific literature, observation and modeling, prompt engineering techniques for efficient text generation using ChatGPT, as well as descriptive and inferential statistical methods for data processing.

To analyze the experimental data, the following statistical procedures were applied: assessment of distribution normality using the Shapiro-Wilk test, Student's t-test for normally distributed data, Wilcoxon signed-rank test and Mann-Whitney U test for non-normally distributed data.

To diagnose the level of critical thinking development, a modified version of Lutsenko's methodology (Lutsenko, 2014) was used and implemented in an online format via the FormsAPP platform. The diagnostic assessment was administered twice: before the introduction of the developed texts and after the completion of the instructional cycles.

To illustrate the structure and content of the modified version of Lutsenko's critical-thinking test used in this study, three sample items are presented below (table 1). In total, the second modified version of the instrument included 14 tasks. Each item was designed not to assess students' subject-matter knowledge in chemistry but rather their ability to evaluate statements, detect inconsistencies, and justify conclusions. Although a chemical context was used to align the tasks with the instructional setting of the experiment, the items did not rely on specific chemical facts, calculations, or definitions. We deliberately attempted to construct the questions in a way that the chemical background would not hinder students' ability to demonstrate critical-thinking skills; however, the chemistry context functioned only as a thematic frame rather than a content requirement.

Table 1 — Sample Items from the Modified Version of Lutsenko's Critical-Thinking Test Used in the Chemistry Context (questions have been translated into English, Correct answers are highlighted in bold)

Item No.	Question	Answer Options
Q1	You are conducting an investigation of how different acids affect metals. What is the most objective way to evaluate the results obtained?	A) Read the theoretical information on the internet. **B) Repeat the experiment several times using different acids and compare the results.** C) Determine the acidity of the solution by tasting it. D) Perform the experiment with only one acid and record the result.

Q2	Which of the following conclusions is logically sound?	A) All solutions of oxygen in water are colorless; therefore, any transparent solution must necessarily contain oxygen. B) Hydrogen is lighter than air; therefore, any gas that is lighter than air contains hydrogen. **C) Two students used an indicator to check the pH of a solution, and both obtained a value of pH < 7. Consequently, the solution is acidic.** D) Magnesium reacts with hydrochloric acid to release hydrogen; therefore, all metals react with acids.
Q8	During a laboratory activity, students carried out the reaction of zinc with hydrochloric acid: $Zn + 2HCl \rightarrow ZnCl_2 + H_2$. Which of the statements below represents a judgment or evaluative opinion rather than a factual statement?	**A) This reaction was the most exciting one. Now I am sure that chemistry is the best subject.** B) A gas was released during the reaction, confirming that a chemical change occurred. C) The reaction of zinc with hydrochloric acid is a single-displacement reaction because a more active metal reacts with the compound. D) The gas released during the reaction can be identified with a burning splint - it will ignite.
Q14	What is incorrect in the following statement? “If a substance is insoluble in water, then it never participates in chemical reactions.”	A) There is nothing incorrect in this statement. **B) It ignores the fact that the substance may react with acids or other reagents.** C) It is correct because all chemical reactions occur in solution. D) It is correct because if a substance does not dissolve, it does not undergo chemical interactions.

These items reflect the overall orientation of the test: identifying logical fallacies, evaluating statements for accuracy, distinguishing between evidence and assumption, and formulating justified conclusions. The chemical elements in the tasks were intended only to situate the questions within the domain of chemistry education while preserving the central focus on critical thinking rather than disciplinary proficiency. The remaining items from the modified version of the test may be provided upon request. Interested readers or reviewers may contact the authors for access to the full set of tasks for academic or research purposes.

The experimental intervention consisted of systematically using educational texts generated by ChatGPT, which, in several places, contained intentional contextual and logical inaccuracies. To create these types of tasks, we used the technique of "prompt engineering," or intelligent query generation. The query generation algorithm consisted of the following: specifying ChatGPT's role and overall goal, and providing an example of one or more texts created by the teacher, allowing the AI to orient itself to the required format. Next, the content itself is provided (the topic, lesson objectives from the lesson plan, the class, and an indication of the inclusion of both content and logical errors). The final step is to provide the AI with feedback on the quality of the generated text and, if necessary, make its own edits. A specific example of a structured prompt (shortened) that was used to generate text with logical and factual errors on the topic "Phosphorus and its compounds": (Act as a "Task Assistant" for a high school Chemistry teacher. Your primary goal is to generate a short, comprehensive educational text (approximately ... words) suitable for 9th-grade students on the topic of Phosphorus and its compounds. The text must be designed as an in-class critical thinking exercise; The text must cover concepts related to: 1. Comparison of phosphorus allotropes (Goal 9.2.1.22). 2. Chemical properties of phosphorus (Goal 9.2.1.23). 3. Locations of phosphorus compound deposits in Kazakhstan (Goal 9.4.2.2); The text MUST contain ... to ... intentionally embedded errors. Include a mix of: Factual Errors: Incorrect chemical data or concepts (e.g., incorrect properties of allotropes, wrong chemical classification). Logical Errors: Inconsistent statements or incorrect causal reasoning (e.g., confusing stability with reactivity, or incorrect localization) *In the first requests, it is also necessary to add a sample text, the format of which the AI will be guided by when creating its own texts.*). Thus, ChatGPT acts as an assistant for generating chemistry-related tasks. Figure 1 illustrates the algorithm used to generate educational chemistry texts containing intentional logical and content errors with the help of ChatGPT. An example of such a text is shown in Figure 2.

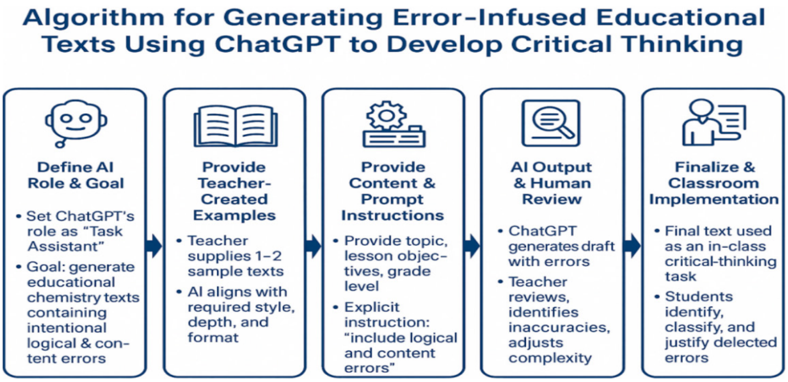
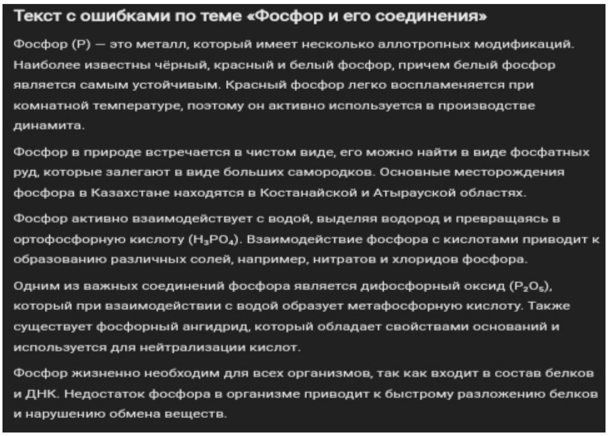


Figure 1 — Algorithm for generating error-infused educational chemistry texts using ChatGPT



Текст с ошибками по теме «Фосфор и его соединения»

Фосфор (P) — это металл, который имеет несколько аллотропных модификаций. Наиболее известны чёрный, красный и белый фосфор, причем белый фосфор является самым устойчивым. Красный фосфор легко воспламеняется при комнатной температуре, поэтому он активно используется в производстве динамита.

Фосфор в природе встречается в чистом виде, его можно найти в виде фосфатных руд, которые залегают в виде больших самородков. Основные месторождения фосфора в Казахстане находятся в Костанайской и Атырауской областях.

Фосфор активно взаимодействует с водой, выделяя водород и превращаясь в ортофосфорную кислоту (H_3PO_4). Взаимодействие фосфора с кислотами приводит к образованию различных солей, например, нитратов и хлоридов фосфора.

Одним из важных соединений фосфора является дифосфорный оксид (P_2O_5), который при взаимодействии с водой образует метафосфорную кислоту. Также существует фосфорный ангидрид, который обладает свойствами оснований и используется для нейтрализации кислот.

Фосфор жизненно необходим для всех организмов, так как входит в состав белков и ДНК. Недостаток фосфора в организме приводит к быстрому разложению белков и нарушению обмена веществ.

Figure 2 — An example of a text on critical thinking

The generated text (Figure 2) serves as a practical case for analysis. It contains several types of errors of varying complexity. For example, a straightforward content error is the misclassification of phosphorus as a metal, which is a fundamental factual mistake contradicting its position in the periodic table. A more subtle and instructive flaw is a logical contradiction embedded in the description of its natural occurrence: the text states that phosphorus is found "in the form of phosphate ores, which occur as large nuggets." This creates an internal inconsistency, as the terms "phosphate ore" (a rock containing chemical compounds of phosphorus) and "nugget" (a native piece of a pure element) are mutually exclusive. This type of error requires the reader to reconcile the meanings of the terms used, moving beyond simple fact-checking to logical analysis.

These text were given to students after covering a specific topic, either at the end of a lesson or at the beginning of the next lesson. Students were given a set amount of time to work with the material (depending on the length of the text). Next, the entire class worked together to identify and classify errors. Students were required not only to identify their errors but also to verbally justify them. This type of work was implemented throughout almost the entire third quarter, with minor modifications (for example, paired work instead of individual work, and error analysis was performed by collecting and reviewing the worksheets at the end of the lesson rather than in person). Such variations allowed the teacher to adjust the level of scaffolding based on students' progress and maintain the novelty of the task format.

Results. *Descriptive Statistics for Pre-test and Post-test Scores.*

Before the implementation of the intervention, all students completed the Lutsenko Critical Thinking Test in order to assess their initial level of critical-thinking skills. Establishing the baseline was essential for evaluating the magnitude of subsequent changes and for verifying that the experimental and control groups did not differ significantly at the outset. The results of the baseline assessment are presented in Table 2.

Table 2 — Results of measuring the initial level of critical thinking

Class	8th Class	9th Class control group	9th Class Experimental Group
Mean value	62.65%	58.28%	59%
Median	62.96%	55.56%	55.56%
Standard deviation	8.66	10.54	9.75

Before introducing error-infused texts for analysis, a short introductory session on critical thinking was conducted in the 8th and experimental 9th grades. The session explained what critical thinking is, why it is important, and how it helps evaluate the accuracy and reliability of information. Students were shown simple examples of logical and factual errors in everyday contexts and were instructed on how to identify, analyze, and justify such errors. This short introduction ensured that all students shared a basic understanding of the skill before beginning the main intervention. This short training was carried out by the same teacher in all classes. After identifying a basic level of critical thinking, the texts were used in the 8th and experimental 9th grades for almost the entire third quarter. These texts were used in almost every lesson (with a few exceptions due to the need for summative assessments and other reasons). In the control 9th grade, the problem-based presentation method was used as an alternative to traditional teaching methods. At the end of the experimental period, the sample of students took the modified Lutsenko test a second time.

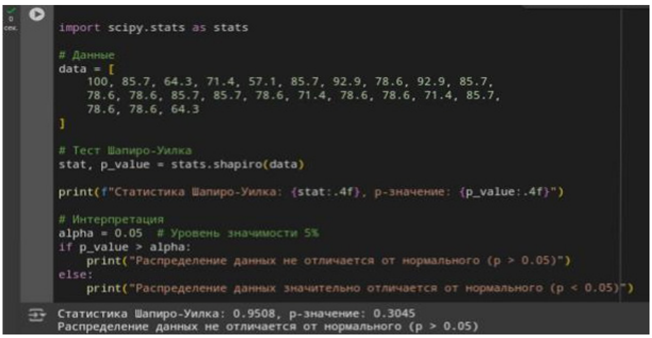
The results of the second critical thinking assessment at the end of the experiment are presented in Table 3.

Table 3 — Results of the second critical thinking test at the end of the experiment

Indicator	8th Class	9th Class control group	9th Class Experimental Group
Mean value	80.2%	63.5%	74.5%
Median	78.6%	78.6%	85.7%
Standard deviation	9.95	29.6	27.4

Normality and Intra-group Analysis.

To identify significant differences between the first test (before using the educational texts for developing critical thinking) and the second test after the experiment, the following algorithm was developed. First, the distribution of values is determined (the Shapiro-Wilk test, example code at figure 3), based on which a comparison criterion is selected (the t-test or Mann-Whitney test). The final step is to determine the p-value. The distribution of test results before and after using texts with logical and chemical errors is presented in Table 4. For the eighth-grade class, the Student's t-test can be used, since the results of both tests are normally distributed (the p-value in both cases is greater than 0.05). For the two ninth-grade classes, the Wilcoxon test was used, since the final test data are not normally distributed (p-value less than 0.05). This approach ensured that the statistical tests applied to each group were consistent with the distributional properties of the data, thereby increasing the reliability of the conclusions regarding the effectiveness of the intervention.



```
import scipy.stats as stats

# Данные
data = [
    100, 85.7, 64.3, 71.4, 57.1, 85.7, 92.9, 78.6, 92.9, 85.7,
    78.6, 78.6, 85.7, 85.7, 78.6, 71.4, 78.6, 78.6, 71.4, 85.7,
    78.6, 78.6, 64.3
]

# Тест Шапиро-Уилка
stat, p_value = stats.shapiro(data)

print(f"Статистика Шапиро-Уилка: {stat:.4f}, p-значение: {p_value:.4f}")

# Интерпретация
alpha = 0.05  # Уровень значимости 5%
if p_value > alpha:
    print("Распределение данных не отличается от нормального (p > 0.05)")
else:
    print("Распределение данных значительно отличается от нормального (p < 0.05)")

Статистика Шапиро-Уилка: 0.9508, p-значение: 0.3045
Распределение данных не отличается от нормального (p > 0.05)
```

Figure 3 — The Shapiro-Wilk criterion program code (Google Colab)

Table 4— Shapiro-Wilk normality test results for student scores before and after the intervention

Class	8th Class	9th Class control group	9th Class Experimental Group
p value before and after the experiment	0.5579 and 0.3045	0.6263 and 0.0206	0.4510 and 0.0016

The results of statistical data processing according to the selected criteria are presented in Table 5. The selection of criteria for identifying statistical significance in Table 5 was determined by the normality of the distribution previously established in Table 4.

Table 5 — Statistical significance of critical thinking score improvements before and after the intervention (according to selected criteria)

Class	8th Class	9th Class control group	9th Class Experimental Group
p value	t-test, $p < 0.001$ (statistically significant difference)	Wilcoxon test 0.4374 (no statistically significant difference)	Wilcoxon test 0.0080 (statistically significant difference)

Analysis of Effect Sizes. While statistical significance testing (p-values) indicates whether the observed changes were likely due to chance, it does not quantify the magnitude of the intervention's impact. To address this, we calculated Cohen’s *d* (Cohen, 1988), a standardized measure of effect size, which allows for an estimation of the practical significance of the pedagogical intervention. Cohen’s *d* was chosen because it provides a unit-less measure of the difference between means expressed in standard deviation units, facilitating comparison across different groups despite variations in baseline scores. Results of Effect Sizes are shown in table 6.

Table 6 — Descriptive statistics and effect sizes for changes in critical thinking scores

Group	*N*	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff (ΔM)	Effect Size (Cohen’s *d*)	Interpretation
8th Experimental	24	62.65 (8.66)	80.20 (9.95)	+17.55	1.88	Very Large Effect
9th Experimental	19	59.00 (9.75)	74.50 (27.40)	+15.50	0.75	Medium-Large Effect
9th Control	15	58.28 (10.54)	63.50 (29.60)	+5.22	0.23	Small Effect

Figure 4 illustrates the mean scores of the post-test across all groups. The error bars represent one standard deviation (SD), highlighting the significant variance in the 9th-grade results (SD = 29.6 and 27.4) compared to the more consistent performance of the 8th-grade experimental group (SD = 9.95).

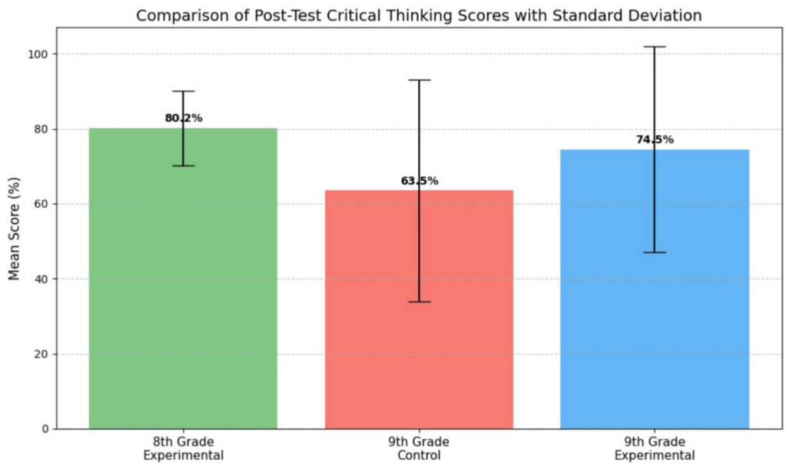


Figure 4 — Mean scores of the post-test across all groups

The dynamics of critical thinking development are illustrated in Figure 5. All groups showed an upward trend in mean scores; however, the intensity of this growth varied significantly. Notably, while the mean scores increased, a significant rise in standard deviation (SD) was observed during the post-test stage, particularly in the 9th-grade groups.

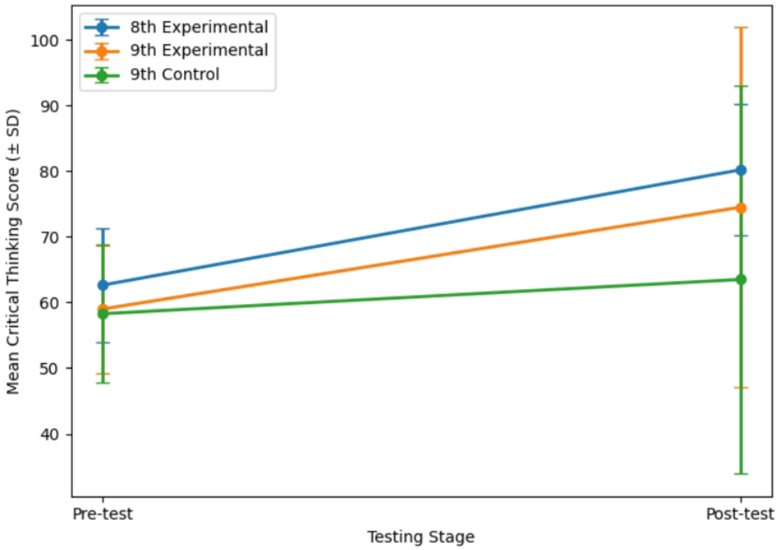


Figure 5 — Pre- and post-test mean critical thinking scores across groups (mean ± SD)

Discussion. An analysis of the primary data (table 2) revealed that the overall sample demonstrated a moderate level of critical thinking. Unfortunately, none of the respondents achieved a high level, exceeding 85% on the percentage scale. It is worth highlighting that the 8th grade class outperformed both 9th grade classes on average. During pedagogical observation during the study period, this class also showed greater interest in the proposed tasks and participated more actively in discussions and group work. Conveniently, both 9th grade classes demonstrated similar results at the beginning of the study. Also, both ninth grades showed comparable standard deviations, indicating closeness and relative uniformity of the samples. This fact will facilitate a more objective interpretation of the impact of the proposed technique at the final stage of the experiment.

Comparing Tables 2 and 3, we can reach the following conclusions. All classes participating in the study improved their average scores compared to the first measurement. The eighth grade maintained this trend, outperforming both ninth grades in average scores. These numerical data align with the observations described above, indicating that students in this class were more open to a relatively new type of task. This confirms the importance of initial internal attitudes, which determine motivation for learning and the application of CT in educational activities. The experimental ninth grade outperformed the control class, but additional statistical tools are required for statistical verification.

To assess changes in critical thinking test scores before and after the intervention, it was necessary to analyze the normality of the distribution of the obtained results (Table 4). The Shapiro-Wilk test for normality of distribution showed that the 8th-grade results both before and after the intervention corresponded to a normal distribution ($p > 0.05$), which allowed the use of parametric statistics (Student's t-test). In contrast, the posttest results of both 9th-grade classes deviated significantly from a normal distribution ($p < 0.05$). This indicates increased heterogeneity in student results after the training period and required the use of nonparametric procedures (Wilcoxon test). The use of statistical methods appropriate to the distribution is necessary for the reliability of the results.

For the 8th and 9th experimental groups, where texts for developing critical thinking were systematically employed as a pedagogical tool, a statistically significant difference was observed between the initial level of critical thinking and after the experiment (table 5). No such pattern was observed in the ninth control grade, where the problem-based method was used (table 5). This indicates that the implemented intervention had a measurable and positive impact on students' ability to detect, analyze, and logically justify inaccuracies in the provided material. Taken together, these findings demonstrate that the deliberate integration of AI-generated texts containing logical and content errors is more effective in stimulating the development of analytical and evaluative reasoning skills than traditional explanatory instruction. The statistically confirmed improvement in both experimental groups supports the hypothesis that targeted interaction with educational texts fosters deeper cognitive engagement and promotes active information processing, which increased the level of critical thinking on average. This pattern also highlights the importance of direct student involvement in error identification and justification -

an element absent in the control condition - indicating that the interactive nature of the intervention may be a key factor in its effectiveness. During the pedagogical observation, it was also discovered that in the experimental groups, as the texts with errors were introduced during full-class discussions, students began to evaluate each other more attentively. Students were also more engaged when conflicting opinions arose regarding the classification of errors as logical or purely chemical. This is consistent with (Andreucci-Annunziata et al., 2023), who indicated that discussions and peer evaluation develop critical thinking.

The Mann-Whitney test was used to compare the final results of the 9th control and experimental classes. Notably, there was no statistically significant difference between the two classes ($p = 0.3852$). This is partly due to the relative similarity of their results, possible design flaws, and other factors (such as guessing). Despite this, given that both classes in which the aforementioned technique was implemented were able to improve their scores on Lutsenko's critical thinking assessment method, the results of the experiment do not fully contradict with our study's hypothesis. It is important to note that this study does not aim to directly compare the absolute critical thinking scores between 8th and 9th-grade students due to inherent age-related cognitive differences and variations in the chemistry curriculum.

As shown in Table 6, the effect sizes reveal distinct patterns across the groups. 8th Grade experimental group: the intervention yielded a very large effect ($d = 1.88$). An effect size of this magnitude indicates a drastic shift in the distribution of scores, suggesting that the "error analysis" method was highly effective for this age group. The students improved by nearly two full standard deviations. 9th Grade Experimental Group demonstrated a medium-to-large effect ($d = 0.75$). Although the raw mean score increased substantially (+15.5%), the effect size was slightly attenuated by the larger standard deviation in the post-test results. Nevertheless, this result confirms a robust positive impact of the intervention. In contrast, the control group showed a negligible-to-small effect ($d = 0.23$). This small value suggests that standard instruction alone resulted in minimal growth in critical thinking skills over the same period. Figure 4 provides a visual comparison of the post-test mean scores across all groups. The experimental classes demonstrate higher average outcomes than the control group, which supports the statistical results.

Particular attention should be paid to the marked increase in the standard deviation among 9th-grade students during the post-test (rising from 9.75 to 27.40 in the experimental group). The visualization in Figure 5 clearly illustrates the «stretching» of the score distribution, indicating a pronounced polarization of outcomes. While a subset of students achieved scores significantly higher than their baseline, others demonstrated a decline, falling below their initial assessment levels. Two primary hypotheses can be proposed to explain this variance. The first is pedagogical: as observed during the intervention, 8th-grade students maintained consistently high levels of engagement, resulting in a stable standard deviation. In contrast, the 9th-grade cohorts exhibited fluctuating levels of interest in the new assignments. This suggests that the intervention's success may be highly dependent on individual intrinsic motivation; highly motivated students were more active and attentive in searching for errors in texts, whereas those with lower engagement may have experienced cognitive overload or a loss of interest. Although various discussion formats (frontal and individual) were implemented to mitigate this, the variance persisted. The second hypothesis is statistical. Given the relatively small sample size in the 9th-grade classes, the results are susceptible to the influence of outliers. In the experimental group, the presence of both maximum (100%) and minimum (21.4%) scores significantly inflated the variance. While this statistical sensitivity is a factor, it underscores the necessity for qualitative inquiry in future studies to explore the underlying causes of such divergent student activity and to refine AI-assisted methodologies for heterogeneous classrooms.

Thus, the results of the experiments implementing educational texts partially support the hypothesis put forward at the beginning of the study. The experimental groups statistically significantly improved their scores on the critical thinking test compared to the initial level. However, the control and experimental 9th-grade students did not show statistically significant differences at the end of the experiment. However, it is worth noting that this study demonstrates the potential of using ChatGPT (and similar chatbots) as a support tool for chemistry teachers. Such AI-powered tools will be most useful when teachers use intelligent queries and prompt engineering. In this case, educational texts will be of the highest quality and will require fewer revisions.

The obtained findings provide answers to the research questions formulated at the beginning of the study. First, the pre-test results confirmed a moderate baseline level of critical thinking. Second, systematic exposure to AI-generated error-infused texts resulted in statistically significant improvements in both experimental groups. Third, the control group showed only minimal progress, indicating that traditional instruction alone was less effective. Finally, the calculated effect sizes demonstrated that the magnitude of the intervention ranged from medium to very large.

While the effect sizes demonstrate the efficacy of the intervention, several limitations of the current study necessitate a cautious interpretation and suggest specific avenues for future inquiry. First, the relatively small sample size and the restriction to specific grade levels limit the generalizability of the findings to broader educational contexts. Second, the notable increase in standard deviation within the 9th-grade experimental group indicates heterogeneity in student responses; while the average performance improved, individual trajectories varied significantly. This suggests that the "error analysis" strategy may impact learners differently depending on their initial cognitive styles or motivation levels.

To address these complexities, we strongly recommend that future research employ mixed-methods designs. While quantitative data confirms that the intervention works, qualitative approaches - such as semi-structured interviews, classroom observations, and "think-aloud" protocols - are essential to understand how it works. Such methods would allow researchers to reconstruct the students' reasoning strategies and identify the specific cognitive mechanisms triggered by AI-generated texts with various types of errors. Specifically, it is crucial to investigate how students detect logical inconsistencies versus factual errors and how they navigate the "cognitive conflict" created by authoritative texts containing deliberate mistakes.

Furthermore, the scope of this pedagogical strategy should be expanded. Future studies should test the scalability of this approach in higher education settings (university level) to determine if error-correction tasks remain effective for learners with more advanced baseline critical thinking skills. Finally, longitudinal studies are needed to assess whether the gains in critical thinking are sustained over time and whether the skills developed through this specific intervention transfer to other academic disciplines.

Conclusion. The study demonstrated that the use of educational texts containing intentional logical and content errors, generated with the support of ChatGPT, constitutes an effective pedagogical technique for fostering students' critical thinking in the context of chemistry education. The quasi-experimental results showed statistically significant improvements in the 8th and 9th grade experimental groups compared to baseline measurements, indicating that the task format encouraged students to evaluate information more critically, identify inconsistencies, and substantiate their judgments.

The findings highlight the potential of integrating AI-based tools into the educational process not merely as sources of content, but as instruments for designing cognitively demanding tasks. Such an approach can enhance learners' analytical engagement and stimulate higher-order thinking skills, provided that AI-generated materials are used intentionally and with methodological control. At the same time, the study underscores the importance of further investigation into the pedagogical boundaries and risks associated with AI integration. Future research should examine the applicability of this technique across different age groups, subjects, and educational levels, explore its long-term impact on learners' cognitive development, and address potential issues related to accuracy, teacher preparedness, and responsible use of AI-generated content in the classroom.

References

Ahern A., et al. (2012) Critical thinking in the university curriculum: the impact on engineering education. European Journal of Engineering Education, 37(2). — P. 125–132. (in English)

Al-Samarraie H., and Hurmuzan S. (2018) A review of brainstorming techniques in higher education. Thinking Skills and Creativity, 27. — P. 78–91. (in English)

Andreucci-Annunziata P., et al. (2023) Conceptualizations and instructional strategies on critical thinking in higher education: A systematic review of systematic reviews. Frontiers in Education, 8. — Article 1141686. (in English)

Butler H.A. (2024) Predicting everyday critical thinking: A review of critical thinking assessments. Journal of Intelligence, 12(2). — P. 16. (in English)

Cohen J. (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum. (in English)

Cruz G., et al. (2021) What critical thinking skills and dispositions do new graduates need for professional life? Views from Portuguese employers in different fields. Higher Education Research and Development, 40(4). — P. 721–737. (in English)

Elshina M.K., et al. (2025) Artificial intelligence in self-learning: new horizons of education. Vestnik Karagandinskogo universiteta. Seriya: Pedagogika, 2(11830). — P. 178–188. (in English)

Hatem R., Simmons B., and Thornton J.E. (2023) Chatbot confabulations are not hallucinations. JAMA Internal Medicine, 183(10). — P. 1177. (in English)

Kim Y., et al. (2023) Towards explainable AI writing assistants for non-native English speakers. arXiv preprint, arXiv:2304.02625. (in English)

Liu Y., and Pásztor A. (2022) Effects of problem-based learning instructional intervention on critical thinking in higher education: A meta-analysis. Thinking Skills and Creativity, 45. — Article 101069. (in English)

Lutsenko E.L. (2014) Adaptatsiya testa kriticheskogo myshleniya L. Starki [Adaptation of the L. Starkey critical thinking test]. Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina. Seriia: Psykholohiia, 55. — P. 65–70. (in Russian)

Raj T., et al. (2022) Importance of critical thinking in education. World Journal of English Language, 12(3). — P. 126–133. (in English)

Rivas S.F., Saiz C., and Ossa C. (2022) Metacognitive strategies and development of critical thinking in higher education. Frontiers in Psychology, 13. — Article 913219. (in English)

Sani S.M., Kurniawati N., and Nurwanti D.I. (2018) The use of reflective writing to improve students' writing and critical thinking skills. (Unpublished manuscript). (in English)

Thornhill-Miller B., et al. (2023) Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. Journal of Intelligence, 11(3). — P. 54. (in English)

Townley A.L. (2018) Teaching and learning science in the 21st century: Challenging critical assumptions in post-secondary science. Education Sciences, 8(1). — P. 12. (in English)

Van Damme D., et al. (2023) Assessing and developing critical-thinking skills in higher education. European Journal of Education, 58(3). — P. 369–386. (in English)

Zahavi H., and Friedman Y. (2020) The Bologna Process: an international higher education regime. In The Bologna Process and its global strategy. — P. 22–38. Routledge. (in English)

Zhai C., Wibowo S., and Li L.D. (2024) The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. Smart Learning Environments, 11(1). — 28. (in English)

Zulyusri Z., et al. (2023) Literature study: Utilization of the PjBL model in science education to improve creativity and critical thinking skills. Jurnal Penelitian Pendidikan IPA, 9(1). — P. 133–143. (in English).

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

For information on Ethics in publishing and Ethical guidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the journals of the Central Asian Academic Research CenterLLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www. elsevier. com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http:// publicationethics. org/files/ u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/ or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:

www: nauka-nanrk.kz

ISSN 2518–1467 (Online),

ISSN 1991–3494 (Print)

http://www.bulletin–science.kz/index.php/en

Managing Editor: A.Shormakova
Editors: D.S. Alenov, M.Konyrbekov
Computer layout: G.D. Zhadyranova

Подписано в печать 27.04.2026.
46,0 п.л.
Заказ 2.

«Central Asian Academic Research Center» LLP
Алматы, Шевченко к-сі, 28