Journal of Health Development
e2663-1776

https://doi.org/10.32921/2663-1776-2025-60-3-jhd009
Short Communication

The need to optimize the assessment of medical students' final knowledge using Artificial intelligence

Received: 02.05.2025
Accepted: 11.06.2025
Published: 30.06.2025

*** Corresponding author:**
Yelena Salikhova
E-mail: Salihova@qmu.kz

Citation: J Health Dev. 2025, 60 (3), jhd009

Xeniya Mkhitaryan [1], Yelena Salikhova [2*], Victor Riklefs [3], Saule Mussabekova [4], Yulia Niyazova [5*], Yelena Mukhametova [6], Bakhyt Suleymenova [7], Kymbat Zhienbayeva [8], Irina Baryshnikova [9]

[1] Head of the Physiology Department, Karaganda Medical University, Karaganda, Kazakhstan
[2] Associate Professor, Karaganda Medical University, Karaganda, Kazakhstan
[3] Acting Chairman of the Board - Rector, Karaganda Medical University, Karaganda, Kazakhstan
[4] Associate Professor, Karaganda Medical University, Karaganda, Kazakhstan
[5] Assistant Professor, Karaganda Medical University, Karaganda, Kazakhstan
[6] Head of the Information Technology Department, Karaganda Medical University, Karaganda, Kazakhstan
[7] Assistant Professor, Karaganda Medical University, Karaganda, Kazakhstan
[8] Assistant Professor, Karaganda Medical University, Karaganda, Kazakhstan
[9] Assistant Professor, Karaganda Medical University, Karaganda, Kazakhstan

Abstract

The Karaganda Medical University has modernized the system for assessing students' final knowledge in the context of modular learning. Testing with one correct answer has been replaced by a system that provides written answers on the session.qmu.kz platform. The platform includes a plagiarism check using the Strikeplagiarizm system. Assessment is carried out according to checklists, including the ability to comment by the examiner.

The aim of this study is to optimize the system for assessing students' final knowledge.

Methods. The study was conducted among teachers and students of the Karaganda Medical University. The sample of students included 358 people from 1st to 5th years who passed exams on the session.qmu.kz platform in the 2022-2023 academic year.

Results. The study identified challenges related to the evaluation of examination papers by faculty and highlighted students' concerns regarding assessment fairness and examiner feedback. These findings support the need for the implementation of artificial intelligence technologies to improve the analysis of responses, the design of examination questions, and the overall evaluation process.

Conclusions. The implementation of an artificial intelligence system in the assessment of the final knowledge of students at medical universities will significantly increase the objectivity of assessments and improve the quality of examination questions, ultimately optimizing the learning process.

Keywords: modular training, session.qmu.kz platform, artificial intelligence, plagiarism check, checklists, objectivity of assessment, optimization of examination questions.

1. Introduction

The problem of objective assessment of students' knowledge in the context of modular training at the Karaganda Medical University (KMU) has become relevant in light of the global challenges of the modern world. It is difficult to imagine modern higher education without digital and information technologies. They allow for flexible organization of the educational process, creation of educational content, synchronous and asynchronous conduct of classes, acquisition of knowledge anywhere in the world, and much more. One of such technologies is artificial intelligence (AI) [1]. The direction under consideration is an important step in improving modern technologies in education and can become the basis for further research in the field of automated assessment of students' knowledge [2,3].

It is undeniable that the role of AI in education is steadily growing, but a number of problems need to be solved to fully realize the benefits of AI in education [4]. In this regard, AI is attractive due to the possibility of increasing efficiency, productivity, saving time and effort, as well as improving overall productivity [5,6].

Until 2019, the approved standard form of passing the exam at the KMU was computer testing using test options with one correct answer, as well as multiple choice tests.

Since 2019, a computerized form of passing the exam in the form of a written answer using a computer keyboard has been launched. For this purpose, the session.qmu.kz platform was developed. In accordance with the thematic plan of the syllabus, an approved list of examination questions of the modular discipline is uploaded to the platform with the ability to select a weighting factor (in %) for each of the disciplines included in the module. The system assumes exam options in the form of a "question-answer", a task or an essay. In order to objectify the assessment, students' examination answers are uploaded to the platform in an impersonal form with the assignment of a unique identification number. For the examiner, the student's work is nameless, having previously undergone a plagiarism check procedure. For this, the Strikeplagiarizm.com system is used, which calculates similarity coefficients, highlights phrases in the answer that are not original, provides links to the used Internet sources, e-books, other student works, resources and knowledge bases identified during the work check.

To score the answer, checklists are uploaded to the platform, according to which the examiner has the opportunity to quantitatively evaluate the student's work from 0 to 100 points. The checklist involves assessing the correctness of the answer, knowledge of theoretical foundations, completeness and logic of what is written, supporting the answer with examples and reasonable conclusions. To justify a decrease in the score, the platform allows you to enter an examiner's comment on the work.

The purpose of the work is to analyze the results of an anonymous survey on the use of the session.qmu.kz platform to improve the system of assessing students' final knowledge in modular training at the KMU. The need to introduce AI as one of the ways to improve the objectivity of assessing examination papers, optimizing the learning process and ensuring the validity of examination questions is considered.

2. Materials and methods

The study was conducted among teachers and students of the KMU. The sample of students included 358 people from 1st to 5th year who passed exams on the session.qmu.kz platform in the 2022-2023 academic year. The average age of respondents was 19.8 years. The survey was also conducted among teachers (n=217), the average age was 42.1 years. A survey on various aspects of the exam was conducted using specially designed questionnaires (2 different questionnaires for the teacher and student) with a 5-point Likert scale. The questionnaires, distributed electronically, consisted of 2 blocks: general and main. The general block included questions about age, gender, educational program in which the respondent is studying or examining, disciplines in which the student took the exam or the teacher checked the exam in the specified year. A special block contained questions about the session.qmu.kz platform using a 5-point Likert scale (from 1 - strongly disagree to 5 - strongly agree) and open questions to provide detailed information. Additionally, the study included data on the time spent by examiners on checking the work, as well as on the frequency of using various exam formats on the platform. Participation in the survey was voluntary. Informed consent was obtained from each respondent, the survey was anonymous.

The obtained data were processed using the statistical software packages Statistica 10.0 and SPSS 20. For categorized data in the responses (1 - "strongly disagree", 2 - "disagree", 3 - "I find it difficult to answer", 4 - "partially agree" and 5 - "strongly agree"), the analysis was carried out with the calculation of proportions and the determination of Wilson 95% confidence intervals for proportions (CI - Confidence interval). To determine differences in age by gender, year of study, and educational program, the Student's t-test was used for quantitative continuous variables and the Pearson $\chi 2$ test for categorical variables.

3. Results

When analyzing the respondents' answers, no statistically significant differences were found by age, gender, year of study, and educational program.

When answering the question about the frequency of using various exam formats on the platform, out of 3 types (essay, task, question-answer), teachers most often preferred the last format 86% (CI [80.13; 90.92]), especially in exams on modular disciplines.

Surveys of examiners showed that 58% (CI [52.35; 64.19]) of respondents' encountered difficulties in processing an excessive number of exam papers during checking, which led to template reviews and dissatisfaction of both students and examiners (Figure 1).

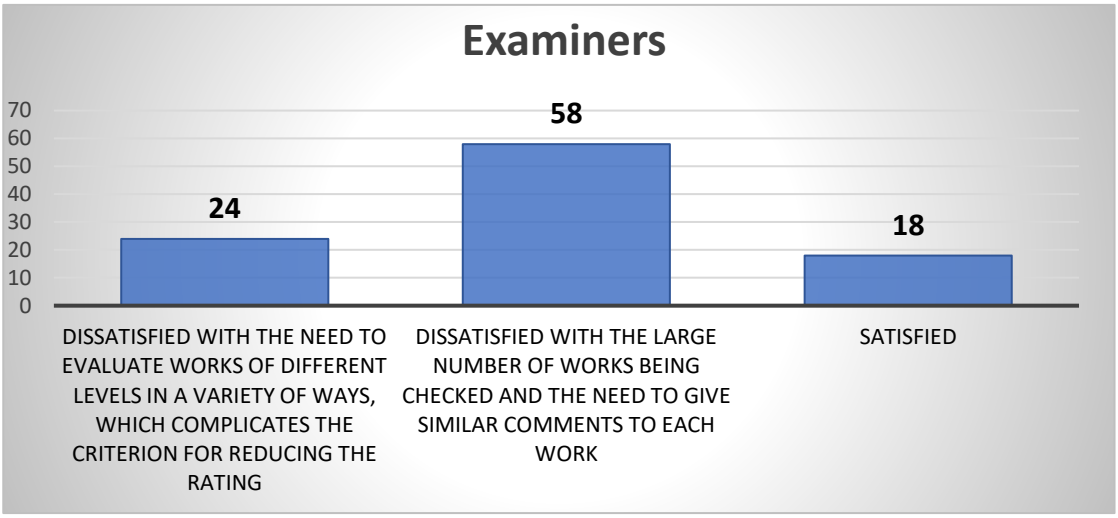

Figure – 1 Dissatisfaction of examiners

The survey results showed that 35% (CI [28.3; 40.7]) of students were dissatisfied with the template feedback, while 15% (CI [13.52; 18.76]) asked for a detailed response to the assessment of the exam paper when appealing. Quite a large proportion of students, 47% (CI [39.1; 53.8]), expressed dissatisfaction with the insufficient time allocated for exams in modular disciplines. In this regard, they had to rationally distribute the total exam time, giving preference to answers to more important exam questions in the module. Students deliberately left questions with a small percentage weight in the overall exam grade for last. As a result, a large proportion of appeals contained a complaint about the lack of time, but in accordance with the Academic Policy of the University, this is not a reason for changing the grade.

The time limit for checking (2 days after the exam) was also the main reason for dissatisfaction indicated in the questionnaire by teachers.

Examiners noted a significant number of similar errors in students' answers, but pointed out the lack of opportunity and time to optimize the assessment of work and improve examination comments. Despite clear and structured checklists, additional analysis was required to understand how many points to reduce the grade by if a student made a particular typical error in his or her answer. At the same time, 25% (CI [21.61; 27.34]) of students complained that different examiners assessed the same errors in their answers differently (Figure 2).

Teachers, in turn, noted the unregulated volume of students' examination papers, which, in their opinion, reduced the effectiveness of checking due to the greater amount of time spent checking voluminous papers.

The checklist was introduced to relieve examiners from the need to write a detailed commentary on each answer. But the number of appeals has not decreased with the introduction of checklists. Appeals require the involvement of three examiners in the appeal

committee with the participation of the dean's office. This significantly reduces overall productivity.

To address these issues, an AI-based decision support system has been proposed.

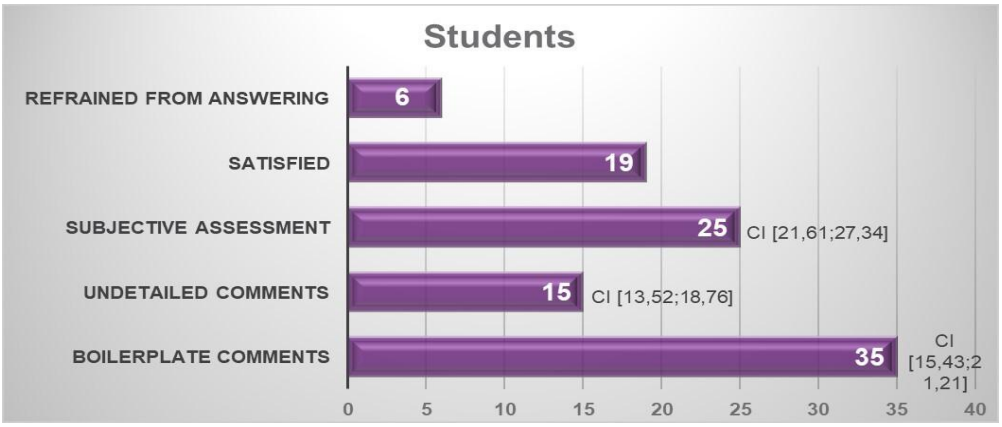

Figure 2 – Student satisfaction

4. Discussion

The AI is supposed to calculate grade reductions based on pre-defined criteria, and the final grades are made jointly by examiners and an appeals committee, taking into account the comments generated by the AI.

To solve these problems, it became necessary to implement an examiner decision support system for the final assessment based on AI. AI is transforming educational assessment by offering innovative solutions for effective and efficient assessment of student learning [7]. AI should calculate grade reduction based on predetermined criteria. An AI-based system can calculate the percentage probability of a desirable or undesirable grade reduction. AI can also be used to identify problems and suggest the same measures to reduce grades for identical errors in different papers, increasing the objectivity of assessors. In controversial cases, final assessments are made jointly by examiners and the appeal committee. It is possible to enhance the impartiality of the decision by taking into account the comments generated by AI. As a result, the use of AI will reasonably lead to an increase in the objectivity of the assessment. In addition, AI can highlight a number of obvious and hidden reasons for unsuccessful answers, students from those questions that none of the examines answer. Or, conversely, identify questions that most students answer easily. This makes it possible to optimize the learning process by paying close attention to the identified gaps in students' knowledge, working through complex issues during practical classes with the use of active learning forms.

Based on the results of checking the examination papers, with a large number of similar errors in answering the question, it becomes necessary to analyze the examination question itself, that is, how valid is the question asked by the teacher, and AI can be used to identify the reasons (misunderstanding of the essence, inaccurate translation or discrepancy with the subject plan, etc.).

Based on the data collected using AI, it becomes possible to analyze the wording of examination questions and determine their validity, as well as optimize the learning process by identifying problem areas and adjusting the teaching and methodological materials and thematic plans accordingly.

5. Conclusion

Thus, the potential use of AI in examination assessment offers significant benefits, but also requires careful development and configuration to ensure maximum efficiency and objectivity of assessment. Implementation of an AI system requires integration with the examination platform, training of staff and consideration of ethical aspects, which is a challenge, but also opens up opportunities for significant improvement of the assessment and training process at KMU.

Conflict of interest. None.

Funding. None.

Author Contributions: Conceptualization – V.R., X.M. and Y.M.; methodology – V.R., X.M.; validation – V.R.; formal analysis – S.M. and Yu.N.; investigation – S.M. and Yu.N.; resources – B.S., K. Zh. and I.B.; data

curation – B.S., K. Zh. and I.B.; writing (original draft preparation) – X.M. and Ye.S.; writing (review and editing) – X.M. and Yu.N.; visualization – X.M. and Ye.S.; supervision – X.M.; project administration – Ye.S. All authors have read and agreed to the published version of the manuscript.

References

1. Buyakova, K., Dmitriev, Ya., Ivanova, A., Feshchenko, A., & Yakovleva, K. (2024). Otnoshenie studentov i prepodavateleĭ k ispol'zovaniyu instrumentov s generativnym iskusstvennym intellektom v vuze (Students' and teachers' attitudes toward the use of generative artificial intelligence tools in universities) [in Russian]. *Obrazovanie i nauka*, 26(7), 160–192. https://doi.org/10.17853/1994-5639-2024-7-160-193
2. Lipina, V. E. (2023). Sistema kontrolya znaniĭ studentov na osnove iskusstvennogo intellekta na urokakh angliĭskogo yazyka (A system for monitoring students' knowledge using artificial intelligence in English language classes) [in Russian]. *Aktual'nye issledovaniya*, 160(30). https://doi.org/10.5281/zenodo.10656471
3. Sysoev, P. V. (2023). Iskusstvennyĭ intellekt v obrazovanii: osvedomlennost', gotovnost' i praktika primeneniya prepodavatelyami vysshehĭ shkoly tekhnologii iskusstvennogo intellekta v professional'noĭ deyatel'nosti (Artificial intelligence in education: Awareness, readiness, and practice of using AI technologies by university faculty in their professional activities) [in Russian]. *Vysshee obrazovanie v Rossii*, 32(10), 9–33. https://doi.org/10.31992/0869-3617-2023-32-10-9-33
4. Tahir, M., Hassan, F. D., & Shagoo, M. R. (2024). Role of artificial intelligence in education: A conceptual review. *World Journal of Advanced Research and Reviews*, 22(1), 1469-1475. https://doi.org/10.30574/wjarr.2024.22.1.1217
5. Ali, O., Abdelbaki, W., Shrestha, A., Elbasi, E., Alryalat, M. A. A., & Dwivedi, Y. K. (2023). A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *Journal of Innovation & Knowledge*, 8(1), 100333. https://doi.org/10.1016/j.jik.2023.100333
6. Flavián, C., & Casaló, L. V. (2021). Artificial intelligence in services: current trends, benefits and challenges. *The Service Industries Journal*, 41(13-14), 853-859. https://doi.org/10.1080/02642069.2021.1989177
7. Kakungulu Samuel, J. (2024). The Role of Artificial Intelligence in Educational Assessment. Available from URL: https://www.researchgate.net/publication/383550444

Жасанды интеллектті қолдану арқылы медицина университетінің студенттерінің қорытынды білімін бағалауды оңтайландырудың қажеттілігі

Мхитарян К.Э. [1], Салихова Е.Ю. [2], Риклефс В.П. [3], Мусабекова С.А. [4], Ниязова Ю.И. [5], Мухаметова Е.Л. [6], Сулейменова Б.М. [7], Жиенбаева К.М. [8], Барышникова И.А. [9]

[1] Физиология кафедрасының меңгерушісі, Қарағанды медицина университеті, Қарағанды, Қазақстан
[2] Қауымдастырылған профессор, Қарағанды медицина университеті, Қарағанды, Қазақстан
[3] Басқарма төрағасының міндетін атқарушы – Ректор, Қарағанды медицина университеті, Қарағанды, Қазақстан
[4] Қауымдастырылған профессор, Қарағанды медицина университеті, Қарағанды, Қазақстан
[5] Ассистент профессор, Қарағанды медицина университеті, Қарағанды, Қазақстан
[6] Ақпараттық технологиялар бөлімінің басшысы, Қарағанды медицина университеті, Қарағанды, Қазақстан
[7] Ассистент профессор, Қарағанды медицина университеті, Қарағанды, Қазақстан
[8] Ассистент профессор, Қарағанды медицина университеті, Қарағанды, Қазақстан
[9] Ассистент профессор, Қарағанды медицина университеті, Қарағанды, Қазақстан

Түйіндеме

Қарағанды медициналық университетінде модульдік оқыту жағдайында студенттердің қорытынды білімін бағалау жүйесін жаңғырту жүргізілді. Бір дұрыс жауапты тестілеуді өткізудің орнына, session.qmu.kz. платформасында жазбаша жауаптарды көздейтін жүйе енгізілді. Платформада Strikeplagiarizm жүйесін

пайдалану арқылы плагиатты тексеру қамтылған. Бағалау емтихан алушының түсініктеме беру мүмкіндігін қоса алғанда, чек-парақтар бойынша жүргізіледі.

Бұл зерттеудің мақсаты - студенттердің қорытынды білімдерін бағалау жүйесін оңтайландыру.

Әдістері. Зерттеу Қарағанды медицина университеті оқытушылары мен студенттері арасында жүргізілді. Студенттердің іріктемесіне session.qmu.kz платформасында 2022-2023 оқу жылында емтихан тапсырған 1-ден 5-ші курсқа дейінгі 358 адам кірді.

Нәтижесі. Оқытушылардың емтихан жұмыстарын бағалаудағы анықталған проблемалар және студенттердің емтихан алушылардың бағалауы мен түсініктемелеріне қанағаттану дәрежесі. Емтихан сұрақтарын талдау, өзгерту және оларды бағалау жүйесі үшін жасанды интеллект енгізуді талап етеді.

Қорытынды. Медициналық университеттер студенттерінің қорытынды білімдерін бағалауға жасанды интеллект жүйесін енгізу бағалаудың объективтілігін едәуір арттыруға және емтихан сұрақтарының сапасын жақсартуға, нәтижесінде оқу процесін оңтайландыруға мүмкіндік береді.

Түйін сөздер: модульдік оқыту, платформа session.qmu.kz, жасанды интеллект, плагиатты тексеру, тексеру парақтары, бағалаудың объективтілігі, емтихан сұрақтарын оңтайландыру.

Необходимость оптимизации оценивания итоговых знаний студентов-медиков с использованием искусственного интеллекта

Мхитарян К.Э. [1], Салихова Е.Ю. [2], Риклефс В.П. [3], Мусабекова С.А. [4], Ниязова Ю.И. [5], Мухаметова Е.Л. [6], Сулейменова Б.М. [7], Жиенбаева К.М. [8], Барышникова И.А. [9]

[1] Заведующая кафедрой физиологии, Карагандинский медицинский университет, Караганда, Казахстан
[2] Ассоциированный профессор, Карагандинский медицинский университет, Караганда, Казахстан
[3] Исполняющий обязанности Председателя Правления – Ректора, Карагандинский медицинский университет, Караганда, Казахстан
[4] Ассоциированный профессор, Карагандинский медицинский университет, Караганда, Казахстан
[5] Ассистент-профессор, Карагандинский медицинский университет, Караганда, Казахстан
[6] Начальник отдела информатизации, Карагандинский медицинский университет, Караганда, Казахстан
[7] Ассистент-профессор, Карагандинский медицинский университет, Караганда, Казахстан
[8] Ассистент-профессор, Карагандинский медицинский университет, Караганда, Казахстан
[9] Ассистент-профессор, Карагандинский медицинский университет, Караганда, Казахстан

Резюме

В Карагандинском медицинском университете проведена модернизация системы оценивания итоговых знаний студентов в условиях модульного обучения. На смену тестированию с одним правильным ответом внедрена система, предусматривающая письменные ответы на платформе session.qmu.kz. Платформа включает проверку на плагиат с использованием системы Strikeplagiarizm. Оценивание производится по чек-листам, включая возможность комментирования экзаменатором.

Целью данного исследования является оптимизация системы оценивания итоговых знаний студентов.

Методы. Исследование проводилось среди преподавателей и студентов Карагандинского медицинского университета. Выборка студентов включала 358 человек с 1 по 5 курс, сдавших экзамены на платформе session.qmu.kz в 2022-2023 учебном году.

Результаты. Выявленные проблемы в оценивании экзаменационных работ преподавателями и степень удовлетворенности студентов оценкой и комментариями экзаменаторов требует внедрения искусственного интеллекта для анализа, модификации экзаменационных вопросов и их системы оценивания.

Выводы. Внедрение системы искусственного интеллекта в оценивание итоговых знаний студентов медицинских университетов позволит значительно повысить объективность оценок и улучшить качество экзаменационных вопросов, в итоге оптимизировать процесс обучения.

Ключевые слова: модульное обучение, платформа session.qmu.kz, искусственный интеллект, проверка на плагиат, чек-листы, объективность оценивания, оптимизация экзаменационных вопросов.