IRSTI 14.35.09
DOI https://doi.org/10.52081/PhSJ.2026.v13.i1.073

A.M. Ivanenko[a]
E-mail: aivanenko940@gmail.com

G.I. Baigunissova[a]
E-mail: baigunisova@mail.ru. *Corresponding author: baigunisova@mail.ru

A.G. Ifatulina[b]
E-mail: alla.ifatulina@gmail.com

[a]*L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*
[b]*Saginov Karaganda Technical University, Karaganda, Kazakhstan*

AUTOMATED WRITING ASSESSMENT IN PRACTICE: AN EXPERIMENTAL ANALYSIS OF AI TOOL FOR WRITING ASSESSMENT

Abstract

This article examines the effectiveness of Artificial Intelligence tools in assessing English speaking and writing skills in contemporary language education. The study explores the theoretical foundations of AI-based assessment, with particular attention to automated writing evaluation and speech recognition technologies. It reviews existing digital platforms used for evaluating productive skills and analyzes their reliability, validity and alignment with CEFR standards. In addition to the theoretical overview, the article presents the results of an experimental study conducted with language learners at intermediate levels. The experiment investigates how AI-generated feedback compares to teacher assessment and evaluates the consistency and accuracy of automated scoring. The findings reveal significant advantages of AI tools, including instant feedback, scalability and scoring consistency while also identifying important limitations in assessing higher-order skills such as coherence, argumentation and communicative effectiveness. The study is concluded by discussing pedagogical implications and recommending a balanced, hybrid assessment approach that combines AI technologies with human expertise.

Keywords:

artificial intelligence, assessment, speaking, writing, feedback.

For citation:

Ivanenko A.M., Baigunissova G.I., Ifatulina A.G. Automated writing assessment in practice: an experimental analysis of ai tool for writing assessment // Philological Sciences Journal. – 2026. – Vol.13. – №1. – Pp. 36-46. DOI https://doi.org/10.52081/PhSJ.2026.v13.i1.073

Introduction

Writing is one of the most complex productive skills in English language learning, requiring mastery of grammar, vocabulary, coherence, organization and style. Traditional teacher-based assessment is valuable, but, limited by time, subjectivity and delayed feedback. With the development of AI, new automated tools can assist both learners and teachers by providing

immediate, personalized and data-driven evaluation. Traditionally, assessment of written work has relied on teachers' professional judgement. This approach is undoubtedly valuable as it enables teachers to consider not only language accuracy, but also meaning, creativity and communicative effectiveness. However, this approach also presents several practical challenges. Checking large numbers of written assignments is time-consuming, feedback can vary depending on the evaluator and students often receive corrections after a delay. When feedback is not immediate, its educational value decreases, and students miss the opportunity to quickly understand and correct their mistakes. In recent years, the rapid development of artificial intelligence has opened up new possibilities for supporting teachers and learners in the writing process. Automated writing evaluation systems can analyse learner texts and provide instant feedback on grammar, vocabulary and structure. These tools are not intended to replace teachers, but rather to support the learning process by providing faster, more accessible and continuous feedback. One example of this technology is the Write & Improve online platform, created by Cambridge English. Students can submit their work and receive automated suggestions for improvement immediately. The growing use of AI tools in language education highlights the importance of understanding how they function in real teaching contexts and the benefits they can offer. While many studies discuss artificial intelligence in general terms, there is still a need for practical analysis of how specific platforms can support writing development, learner independence and formative assessment in everyday classroom practice.

This article's novelty focus is on the pedagogical use of AI-based writing evaluation as a supportive instructional tool, rather than simply a technological innovation. Particular attention is given to how automated feedback can help learners identify their mistakes, revise their texts more effectively, and take a more active role in their own learning process.

This article examines the role of AI-based writing assessment tools in English language learning, paying particular attention to the possibilities offered by the Write & Improve platform. The study explores how automated evaluation can complement traditional teacher feedback, contributing to a more efficient, timely and learner-centred approach to writing instruction.

Materials and research methods

During the study the experiment was conducted to investigate the effectiveness of AI-based automated writing feedback for learning English. The experiment involved two groups of English learners: a control group and an experimental group. Both groups studied under comparable instructional conditions and followed the same writing syllabus. However, different feedback approaches were applied during the assessment stage.

The study participants were A2-B1 level English learners. They were divided into two groups of approximately equal size. Prior to the experiment, both groups completed a diagnostic writing task to determine their initial writing proficiency and ensure comparability between the groups. During the experimental period, both groups were given the same writing tasks, including short essays, opinion paragraphs and descriptive texts, all of which were aligned with standard communicative learning objectives. The control group received traditional teacher-based feedback consisting of written corrections, comments and recommendations, which were provided after the assignments had been submitted.

In contrast, the experimental group used Write & Improve, an automated writing evaluation platform developed by Cambridge English. Students submitted their work via the system and received immediate automated feedback, including highlighted errors, suggestions for improving their language use, and an estimated proficiency level. They were encouraged to revise and resubmit their texts multiple times based on the system’s recommendations, thereby engaging in an iterative self-correction process. The experimental instruction period lasted for 1 week (Three lessons in each group). At the end of the experimental period, both groups completed a final writing task of a similar difficulty to the initial diagnostic assignment. The collected texts were evaluated

according to unified assessment criteria, including grammatical accuracy, lexical diversity, coherence and task achievement.

A comparative method was applied to analyse the results and examine differences in writing performance between the control and experimental groups. Quantitative analysis compared average scores before and after the intervention, while qualitative observation focused on patterns of error reduction, revision behaviour and learner engagement with feedback.

Thus, the study's methodological framework combined experimental comparison, quantitative score analysis and qualitative evaluation of learner writing development to determine the effectiveness of AI-supported automated writing assessment as a pedagogical tool.

Literature Review

Research on AI-assisted writing assessment shows several important features of Automated Writing Evaluation (AWE) systems. These systems are designed to provide immediate feedback, spot grammatical and lexical errors, evaluate how well the text flows together, and generate CEFR-aligned scoring. Their development is based on two things: large-scale collection of language data and machine learning technologies. These technologies can automatically recognise patterns in learner texts that have already been scored (Li, 2020). From a technical point of view, AWE systems use supervised learning models that are trained on human-rated corpora. This means they can approximate human scoring in controlled conditions (Deane, 2022).

A key idea in the theoretical discussion of AI-based assessment is construct validity. This is how well a tool measures the intended language ability. In language testing theory, validity means that assessments have to show real communication skills, not just single language features (Alderson, 2021). However, scholars say that automated systems tend to focus on measurable surface-level features such as grammar, vocabulary range, and sentence complexity. This may lead to some things being missed (Deane, 2022). While AI scoring is quite reliable at evaluating how well language is used (Li, 2020), it is less good at assessing things like creativity, how well arguments are made, and how suitable language is for the situation (Zhai, 2022).

One well-known AWE platform is Write & Improve, which was developed by Cambridge Assessment English. The system has been trained on a large amount of learner data and provides instant feedback that is aligned with the Common European Framework of Reference for Languages (CEFR). This feedback is designed to support learning in order to improve your language skills (Cambridge Assessment English, 2020). It has been used in schools and universities for testing, students' own practice and getting ready for exams. The system has a clear scoring system and gives immediate feedback, which encourages people to make changes and learn independently. However, research consistently highlights that AI tools should support teachers, rather than replace them (OECD, 2021).

Another concern is about fairness and bias. AI systems need the data used to train them, and imbalances in the data may change the results (UNESCO, 2021). Additionally, automated evaluation systems may find it hard to take into account cultural differences in communication or effective writing strategies that are not common (Zawacki-Richter et al., 2022). These limitations show how important it is for teachers to be there to make sure that learning is ethical and that the students are being taught well.

Overall, theoretical framework indicates that AI-assisted writing assessment is highly effective in delivering fast, consistent, and scalable feedback, particularly with regard to linguistic accuracy. However, its limitations in evaluating complex communicative competence justify the implementation of a hybrid assessment model that combines technological efficiency with professional human judgment (Lee, 2022). This theoretical foundation provides the conceptual underpinnings for the empirical component of the present study, which employs experimental methods to investigate the reliability and pedagogical value of AI-based writing assessment in comparison with teacher evaluation.

Discussion and results

Many digital tools now help learners analyze their texts, receive targeted feedback and improve their writing through repeated revision. As Cambridge Assessment notes, AI supported evaluation “can provide fast, consistent feedback to learners while supporting teachers in managing large amounts of written work” (Cambridge English, 2020) Nowadays, there is a wide choice among various AI-based writing assessment Tools divided into categories:

1. Grammar and sentence-focused tools,
2. Vocabulary enhancement tools,
3. Plagiarism-detection systems,
4. Holistic scoring engines,
5. Comprehensive Automated Writing Evaluation (AWE) systems

Our investigation is lying into the 5th category (AWE systems) and one of the tools that are focused on providing full-scale analysis, feedback and scoring is Write&Improve. Cambridge English describes it as a platform that “automatically evaluates learner’s writing and provides immediate feedback for improvement” (Cambridge English, 2021)

Reliability refers to how consistent and stable the scoring is. “Artificial intelligence is reshaping educational assessment by enabling scalable, data-driven and adaptive feedback mechanisms.” (Zhai, 2022) According to the Write&Improve help documentation, the system relies on supervised machine learning trained on the Cambridge Learner Corpus, which includes millions of learner essays that were previously scored by human examiners. “…a database of 30 million words of EFL learners’ essays…along with the scores they awarded by human examiners” This use of large-scale, human-annotated data contributes to the internal consistency of the scoring engine. Moreover, the system’s design is deliberately conservative: it “only indicated possible errors when it’s more than 90% certain it’s right”. This strategy reduces the risk of false positives in error marking, which further strengthens reliability. However, there are inherent limitations to validity. “AI in language assessment presents both unprecedented opportunities for efficiency and serious challenges regarding validity, transparency and fairness.” (Lee, 2022) Automatic feedback tends to excel in detecting surface-level issues (grammar, spelling, lexical errors), but may struggle with higher-order writing skills such as argument quality, cohesion or rhetorical sophistications. Another question that may arise is about ethics. “AI systems in education must be human-centered and aligned with pedagogical values.” (UNESCO, 2021) Therefore, although Write&Improve provides reliable, CEFR-based feedback and is valid for many formative purposes, its validity is ultimately constrained by its inability to fully replicate deeper human judgement. To maximize its effectiveness, it should be used in conjunction with teacher feedback or as a part of a hybrid assessment model. “The integration of AI into assessment systems should support, not replace professional judgment.” (OECD, 2021)

“Despite rapid technological development, AI applications in education remain limited in their capacity to interpret complex human communication.” (Zawacki-Richter et al., 2022) Artificial intelligence is becoming more and more important in how student writing is evaluated. This offers new ways to provide feedback and scoring on a large scale. But AI-based writing tools also have big limitations that must be considered when using them in teaching. “The key challenge for AI in language testing is ensuring that automated scores genuinely represent the intended construct.” (Alderson, 2021).

One of the best things about using AI to evaluate writing is how fast it is and how well it can be used on a large scale. “Automated scoring systems may overemphasize easily quantifiable features at the expense of deeper discourse qualities” (Deane, 2022) Automated tools can analyse a text very quickly, which makes them particularly useful in educational settings with a lot of students. Cambridge English says that combining automated and human marking can "deliver faster results with the same level of accuracy as fully human-marked exams." This shows how AI can speed up the assessment cycle without affecting the overall quality of hybrid systems. Another good thing about it is that it is consistent. Unlike people, who may have different ideas about how to

score things, or get tired, AI systems give all learners the same score. For example, the 'Write & Improve' tool uses an artificial intelligence system that has been trained using thousands of essays from the Cambridge Learner Corpus. This means it can give consistent scores for different writing samples. AI-based tools also give students immediate feedback, which helps them learn by themselves. Apps like Write & Improve point out mistakes in grammar and vocabulary and suggest ways to improve your text. This lets you write, edit and resubmit your texts whenever you like. This approach helps students focus on areas they may not be able to identify on their own, encouraging them to take more control of their learning. Finally, AI feedback is easy to get and cheap, especially for schools that don't have much money. Many of these platforms are free or low-cost, which makes things easier for teachers and gives students more chances to practise writing. But there are important limits to these tools. First, they tend to focus on basic mistakes, such as spelling, grammar, and choosing the right words. While this can be useful for accuracy, automated systems often struggle with complex rhetorical features such as coherence, cohesion, argument structure, and creativity.

As the Write & Improve instructions explain, the system cannot judge how well the essay answers the question, showing that current AI is not yet able to assess content relevance and higher-order thinking. Another problem is that it might not be completely accurate. Although tools like Write & Improve are based on large sets of data and are usually reliable, they are not perfect. If learners are only given automated feedback and don't get any feedback from teachers, they might get the wrong idea about what they're doing. This is because there might be some mistakes in the automated feedback, like sometimes getting a positive when a negative would be better, or sometimes getting a negative when a positive would be better.

Also, AI systems do not always understand cultural differences or the right tone and register. This can be a real problem when you're doing writing tasks that need you to think about how things actually work in the real world, or when you need to use language that fits with the situation you're writing about. Another worry is that students will start to rely too much on AI, which might stop them from thinking critically about their own writing. When feedback is instant and automatic, learners may accept suggestions without understanding why a particular change is necessary. Finally, there are ethical considerations, including data privacy, bias in training corpora, and questions about transparency. Machine-learning tools are based on the data they are built on. If the data includes biases related to things like learner background or writing style, then some students may be unfairly penalised. Combined assessment may provide even better effect on learners' results. "Emeging AI tools such as ChatGPT are transforming how learners receive feedback, but their pedagogical integration requires critical oversight" (Kohnke, Moorhouse and Zou, 2023). Overall, AI-based writing evaluation tools are incredibly useful for both learners and teachers, offering immediate feedback, consistency, and efficiency. But they shouldn't be seen as a replacement for human judgment. The best approach combines the strengths of AI – speed, scalability and surface-level accuracy – with the deeper analytical and pedagogical insights provided by human teachers.

The experiment was done over a week in two classes of intermediate-level English learners. These classes were on at different times of the day, followed different schedules, and weren't linked to each other academically. This made sure that the groups didn't influence each other's performance or writing habits. A total of 17 students took part:

- There are 10 students in the control group (Class A).
- The experimental group is made up of 7 students (Class B).

Both groups were pretty similar in their general English proficiency and writing experience, which made the comparison a decent idea.

The experiment tried to find out if using an AI-based writing assessment tool (Write & Improve) could improve students' writing in a short space of time. In brief, Write&Improve is a tool that was developed by Cambridge, and is widely used all around the world. It lets thousands of students make their writing more proficient and confident. The first step was to present this tool to

my experimental group, so they could know what they are dealing with. After short and flexible registration we proceed to two different interfaces: for teacher and for students.

The experiment aimed to investigate whether using an AI-based writing assessment tool (Write&Improve) could improve students' writing within a short period. The focus was placed on three writing components:

- Grammatical accuracy,
- Vocabulary range and appropriateness,
- Coherence and overall quality.

Both groups used the same writing prompts and the same CEFR-based rubric. The only difference was the feedback method:

- Control Group: traditional teacher feedback only, no AI tools.
- Experimental Group: teacher guidance + Write&Improve for automated assessment, error checking and iterative rewriting.

Organisation of the Experiment. The two groups had their classes at different times, so lessons were held separately. The tasks were identical, but delivered their regular class hours. The experiment included three writing sessions across one week (Table1, Table 2).

Table 1 – Control Group Results

Student	Baseline Score	Final Score	Progress	Group
A1	B1.1	B1.2	+0.1	Control
A2	B1.1	B1.1	0	Control
A3	B1.2	B1.2	0	Control
A4	B1.1	B1.2	+0.1	Control
A5	B1.2	B1.3	+0.1	Control
A6	B1.1	B1.2	+0.1	Control
A7	B1.2	B1.2	0	Control
A8	B1.1	B1.2	+0.1	Control
A9	B1.2	B1.2	0	Control
A10	B1.1	B1.2	+0.1	Control
Average Progress	—	—	**+0.07**	Control

Table 2 – Experimental Group Result

Student	Baseline Score	Final Score	Progress	Group
B1	B1.1	B1.2	+0.1	Experimental
B2	B1.1	B1.3	+0.2	Experimental
B3	B1.2	B1.3	+0.1	Experimental
B4	B1.1	B1.2	+0.1	Experimental
B5	B1.2	B1.3	+0.1	Experimental
B6	B1.1	B1.3	+0.2	Experimental
B7	B1.2	B1.3	+0.1	Experimental
Average Progress	—	—	**+0.13**	Experimental

The results of the mini-experiment demonstrate measurable differences in writing progress between the control and experimental groups. As shown in Table 1, the feedback showed only minor improvement in their writing skills during the study. While some students improved slightly (for example, from B1.1 to B1.2), others showed no observable change. The calculated average progress for the control group was +0.07, indicating relatively limited development during the instructional period. By contrast, the experimental group demonstrated more consistent and noticeable improvement (see Table 2).One possible explanation for this outcome is the immediacy and iterative nature of the feedback provided by the Write & Improve platform, which was developed by Cambridge English. Unlike traditional delayed correction, this system allowed students to instantly identify language errors, revise their texts and resubmit improved versions multiple times. This repeated interaction with feedback probably helped learners to become more aware of their errors and encouraged them to engage more actively in the revision process.

However, it should be noted that the experiment was relatively short (three days), which may have limited the extent of measurable improvement. Nevertheless, even within this short period, the experimental group demonstrated stronger positive dynamics than the control group, suggesting that AI-assisted formative feedback can enhance short-term writing development and learners' responsiveness to correction. Overall, the comparative analysis of the two tables confirms that integrating automated writing evaluation tools can positively influence students' writing progress when used alongside traditional teaching methods as a supplementary instructional resource.

Conclusion

The results of the experiment suggest that automated assessment could be a valuable addition to the evaluation of productive language skills, particularly in formative assessment contexts where regular practice and prompt feedback are important. This study examined the effectiveness of Write & Improve, an AI-based platform developed by Cambridge English, in evaluating the writing performance of EFL learners at proficiency levels ranging from A2 to B2.

The findings show that the platform can provide immediate, structured feedback, enabling learners to swiftly identify and correct linguistic errors. In particular, the system proved effective in identifying common grammatical errors, lexical inaccuracies, and basic structural issues. The ability to receive instant responses and repeatedly revise written work encouraged learners to engage more actively in the editing process, supporting the development of self-correction strategies. From a pedagogical perspective, the immediacy of this feedback is one of the most significant advantages

of AI-assisted writing assessment, as it helps to maintain learner motivation and promote continuous practice.

However, the study also revealed limitations of automated evaluation. While the system reliably identifies surface-level language errors, it does not consistently capture more complex aspects of writing quality, such as logical coherence, strength of argumentation, stylistic appropriateness and overall communicative effectiveness. These higher-level components remain strongly dependent on contextual interpretation and professional pedagogical judgement. Consequently, the findings suggest that AI-based tools should be viewed as a complementary resource that can enhance the overall feedback process, rather than as a replacement for teacher assessment.In conclusion, the present study confirms the growing practical potential of artificial intelligence in modern language education. At the same time, however, it emphasises that effective writing assessment requires a balanced combination of technological assistance and human expertise. The most productive instructional model appears to be one in which automated systems provide rapid diagnostic support while teachers continue to guide learners in developing deeper communicative and academic writing skills.

References

Alderson, J. Charles (2021) ‘Artificial intelligence and language testing: The challenge of validity’, Language Testing, 38(3), pp. 1–15.

Cambridge Assessment English (2020) The future of automated assessment and human marking. Cambridge: Cambridge University Press.

Deane, Paul (2022) ‘Automated scoring and the problem of construct representation’, Assessing Writing, 52, 100609.

Kohnke, Lucas, Moorhouse, Benjamin Luke and Zou, Di (2023) ‘ChatGPT for language teaching and learning’, RELC Journal, 54(2), pp. 1–14.

Lee, John S. (2022) ‘Artificial intelligence in language assessment: Opportunities and challenges’, Language Testing in Asia, 12(1), pp. 1–17.

Li, Jin (2020) ‘Artificial intelligence in automated writing evaluation: Reliability and validity considerations’, Journal of Second Language Writing, 49, pp. 1–12.

Mayfield, Elijah and Black, Alan W. (2021) ‘Bias and fairness in automated speech scoring’, Proceedings of the ACL Workshop on Educational NLP, pp. 45–54.

OECD (2021) AI in Education: Guidance for Policy Makers. Paris: OECD Publishing.

UNESCO (2021) AI and Education: Guidance for Policy-Makers. Paris: UNESCO.

Zawacki-Richter, Olaf, Marin, Victoria I., Bond, Melissa and Gouverneur, Franziska (2022) ‘Systematic review of research on artificial intelligence applications in higher education’, International Journal of Educational Technology in Higher Education, 19(1), pp. 1–27.

Zhai, Xiaoming (2022) ‘Artificial intelligence for education: A review’, Education and Information Technologies, 27, pp. 1–19.

А.М. Иваненко[a]
E-mail: aivanenko940@gmail.com

Г.И. Байгунисова[a]
E-mail: baigunisova@mail.ru. *Корреспондент авторы: baigunisova@mail.ru

А.Г. Ифатулина[b]
E-mail: alla.ifatulina@gmail.com

[a]*Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан*
[b]*А. Сағынов атындағы Қарағанды техникалық университеті, Қарағанды, Қазақстан*

ТӘЖІРИБЕДЕ ЖАЗБАША СӨЙЛЕУДІ АВТОМАТТАНДЫРЫЛҒАН БАҒАЛАУ: ЖАЗБАША СӨЙЛЕУДІ БАҒАЛАУҒА АРНАЛҒАН AI ҚҰРАЛЫН ЭКСПЕРИМЕНТТІК ТАЛДАУ

Аннотация. Бұл мақалада қазіргі тілдік білім беру жүйесінде ағылшын тіліндегі ауызша және жазбаша сөйлеу дағдыларын бағалауда Жасанды интеллект (ЖИ) құралдарының тиімділігі қарастырылады. Зерттеу ЖИ негізіндегі бағалаудың теориялық негіздерін талдайды, әсіресе жазбаша жұмыстарды автоматты бағалау жүйелері мен сөйлеуді тану технологияларына ерекше назар аударады. Мақалада продуктивті дағдыларды бағалауға арналған қолданыстағы цифрлық платформаларға шолу жасалып, олардың сенімділігі, валидтілігі және CEFR стандарттарына сәйкестігі талданады. Теориялық талдаумен қатар, мақалада орта деңгейдегі тіл үйренушілер арасында жүргізілген эксперименттік зерттеу нәтижелері ұсынылады. Эксперимент ЖИ арқылы берілген кері байланысты мұғалім бағалауымен салыстыруға, сондай-ақ автоматтандырылған бағалаудың тұрақтылығы мен дәлдігін анықтауға бағытталған. Нәтижелер ЖИ құралдарының жедел кері байланыс беруі, ауқымды қолдануға бейімділігі және бағалаудың бірізділігі сияқты маңызды артықшылықтарын көрсетеді, сонымен бірге мәтіннің байланыстылығы, аргументациясы және коммуникативтік тиімділігі сияқты жоғары деңгейлі дағдыларды бағалауда белгілі бір шектеулер бар екенін айқындайды. Зерттеу қорытындысында педагогикалық маңыздылығы талқыланып, ЖИ технологиялары мен мұғалімнің кәсіби бағалауын үйлестіретін теңгерімді гибридті тәсілді қолдану ұсынылады.

Тірек сөздер: жасанды интеллект, бағалау, ауызша сөйлеу, жазбаша сөйлеу, кері байланыс.

Сілтеме жасау үшін: Иваненко А.М., Байгунисова Г.И., Ифатулина А.Г. Тәжірибеде жазбаша сөйлеуді автоматтандырылған бағалау: жазбаша сөйлеуді бағалауға арналған AI құралын эксперименттік талдау // Philological Sciences Journal. – 2026. – Vol.13. – №1. – 36-46. DOI https://doi.org/10.52081/PhSJ.2026.v13.i1.073

А.М. Иваненко[a]
E-mail: aivanenko940@gmail.com

Г.И. Байгунисова[a]
E-mail: baigunisova@mail.ru. *Автор для корреспонденции: baigunisova@mail.ru

А.Г. Ифатулина[b]
E-mail: alla.ifatulina@gmail.com

[a]*Евразийский национальный университет имени Л.Н.Гумилева, Астана, Казахстан*
[b]*Карагандинский технический университет имени А. Сагинова, Караганда, Казахстан*

АВТОМАТИЗИРОВАННАЯ ОЦЕНКА ПИСЬМЕННОЙ РЕЧИ НА ПРАКТИКЕ: ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ ИИ-ИНСТРУМЕНТА ДЛЯ ОЦЕНКИ ПИСЬМЕННОЙ РЕЧИ

Аннотация. В данной статье рассматривается эффективность инструментов искусственного интеллекта (ИИ) при оценивании навыков устной и письменной речи на английском языке в современном языковом образовании. Исследование анализирует теоретические основы ИИ-ориентированного оценивания с особым вниманием к системам автоматической проверки письменных работ и технологиям распознавания речи. В работе представлен обзор существующих цифровых платформ, используемых для оценки продуктивных речевых навыков, а также анализируются их надёжность, валидность и соответствие стандартам CEFR. Помимо теоретического обзора, статья включает результаты экспериментального исследования, проведённого среди изучающих язык на среднем уровне. Эксперимент направлен на сравнение обратной связи, сгенерированной ИИ, с оцениванием преподавателя, а также на определение согласованности и точности автоматического оценивания. Полученные результаты выявляют значительные преимущества ИИ-инструментов, включая мгновенную обратную связь, масштабируемость и стабильность оценивания, а также указывают на существенные ограничения при оценке навыков более высокого уровня, таких как связность текста, аргументация и коммуникативная эффективность. В заключение обсуждаются педагогические импликации и рекомендуется сбалансированный гибридный подход, сочетающий ИИ-технологии и экспертную оценку преподавателя.

Ключевые слова: искусственный интеллект, оценивание, устная речь, письменная речь, обратная связь.

Для цитирования: Иваненко А.М., Байгунисова Г.И., Ифатулина А.Г. Автоматизированная оценка письменной речи на практике: экспериментальный анализ ИИ-инструмента для оценки письменной речи // Philological Sciences Journal. – 2026. – Vol. 13. – №1. – С. 36-46. DOI https://doi.org/10.52081/PhSJ.2026.v13.i1.073

Information about authors:

Ivanenko A.M., master student of the Foreign Languages Theory and Practice Department. L.N. Gumilyov Eurasian National University, Satbayev str. 2, Astana, Kazakhstan.
ORCID: 0009-0007-1274-5076

Baigunissova G.I., Doctor of philology, dotzent, Professor of the Foreign Languages Theory and Practice Department, L.N. Gumilyov Eurasian National University, Satbayev str. 2, Astana, Kazakhstan.
ORCID: 0000-0001-8964-853X

Ifatulina A.G., Senior teacher of Foriegn Languages Department. Saginov Karaganda Technical University, Karaganda, Kazakhstan.
ORCID 0009-0000-7955-3238

Авторлар туралы мәлімет:
Иваненко Алёна Максимовна, магистрант, Л.Н. Гумилев атындағы Еуразия үлттық университеті, Сәтбаев көш., 2, Астана, Қазақстан.
ORCID: 0009-0007-1274-5076

Байгунисова Гүлбаршын Ыдырысқызы, филология ғылымдарының докторы, доцент. Шетел тілдерінің теориясы және тәжірібесі кафедрасының профессоры, Л.Н. Гумилёв атындағы Еуразия ұлттық университеті, Астана, Қазақстан.
ORCID: 0000-0001-8964-853X

Ифатулина Алла Николаевна, А.Сағынов атындағы Қарағанды техникалық университетінің шетел тілдері кафедрасының аға оқытушысы, Қарағанды қ., Қазақстан.
ORCID: 0009-0000-7955-3238

Сведения об авторах:
Иваненко Алёна Максимовна, магистрант, Евразийский национальный университет, ул. Сатпаева 2, Астана, Казахстан**.**
ORCID: 0009-0007-1274-5076

Байгунисова Гульбаршин Идрисовна, доктор филологических наук, доцент, профессор кафедры теории и практики иностранных языков. Евразийский национальный университет им. Л.Н. Гумилева. Астана, Казахстан.
ORCID ID: 0000-0001-8964-853X

Ифатулина Алла Николаевна, старший преподаватель кафедры иностранных языков. Карагандинский технический университет имени А.Сагинова, Караганда, Казахстан.
ORCID: 0009-0000-7955-3238

The article was submitted on 15.03.2026;
approved after reviewing on 02.04.2026; accepted for publication on 05.04.2026.
Мақала редакцияға 15.03.2026 ж. келіп түсті;
02.04.2026 ж. рецензиядан кейін мақұлданды; 10.04.2026 ж. баспаға қабылданды.
Статья поступила в редакцию 15.03.2026 г.;
одобрена после рецензирования 02.04.2026 г.; принята к публикации 05.04.2026.