INTERNATIONAL MULTIDISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT
SJIF 2019: 5.222 2020: 5.552 2021: 5.637 2022:5.479 2023:6.563 2024: 7,805
eISSN :2394-6334 https://www.ijmrd.in/index.php/imjrd Volume 12, issue 05 (2025)

THE ROLE OF AI AND AUTOMATED SCORING IN ENGLISH LANGUAGE PROFICIENCY TESTS

Mardonova Gulchehkra Vasliddin kizi

Teacher of English Philology Department,

Jizzakh State Pedagogical University.

E-mail: mardanova-95@list.ru

Annotation: This paper explores the role of artificial intelligence (AI) and automated scoring systems in English proficiency tests. AI technologies are not only being used to create test assignments, but also to analyze and evaluate responses. Automated scoring systems serve to reduce human errors, speed up the grading process, and increase fairness. The article also analyzes the advantages and disadvantages of AI-based systems, their impact on the reliability of test results, and prospects for evaluating English tests in the future.

Key words: artificial intelligence (ai), automated scoring, English proficiency test, assessment process, test reliability, ai technology, student outcome analysis, language test automation, digital assessment systems, fair assessment of tests.

Аннотация: В этой статье исследуется роль искусственного интеллекта (ИИ) и автоматизированных систем подсчета баллов в тестах на знание английского языка. Технологии искусственного интеллекта используются не только для создания тестовых заданий, но и для анализа и оценки ответов. Автоматизированные системы подсчета баллов позволяют сократить количество человеческих ошибок, ускорить процесс выставления оценок и повысить справедливость. В статье также анализируются преимущества и недостатки систем на основе искусственного интеллекта, их влияние на достоверность результатов тестов и перспективы оценки тестов по английскому языку в будущем.

Ключевые слова: искусственный интеллект (ии), автоматизированный подсчет баллов, тест на знание английского языка, процесс оценивания, надежность тестов, технология искусственного интеллекта, анализ результатов учащихся, автоматизация языковых тестов, цифровые системы оценивания, справедливое оценивание тестов.

Annotatsiya: Ushbu maqola ingliz tilini bilish testlarida sun’iy intellekt (AI) va avtomatlashtirilgan ball berish tizimlarining rolini o‘rganadi. AI texnologiyalari nafaqat test topshiriqlarini yaratish, balki javoblarni tahlil qilish va baholashda ham foydalanilmoqda. Avtomatlashtirilgan ball berish tizimlari inson xatolarini kamaytirish, baholash jarayonini tezlashtirish va adolatni oshirishga xizmat qilmoqda. Shuningdek, maqolada AI asosida ishlaydigan tizimlarning afzalliklari va kamchiliklari, ularning test natijalarining ishonchliligiga ta’siri, va kelajakda ingliz tili testlarini baholashdagi istiqbollari tahlil qilinadi.

Kalit so‘zlar: sun’iy intellekt (ai), avtomatlashtirilgan ball berish, ingliz tilini bilish testi, baholash jarayoni, test ishonchliligi, ai texnologiyalari, o‘quvchilarning natijalarini tahlil qilish, til testlarining avtomatizatsiyasi, raqamli baholash tizimlari, testlarni adolatli baholash.

INTRODUCTION

Digital assessment has become increasingly popular in recent years. But what is the role of digital assessment in education today? We want to give you a little insight into digital assessment and automated scoring.

A few years ago, there might have been doubts about the role of AI in English language assessment and the ability of computers to accurately score language tests. But today, thousands of teachers around the world use automated language tests to assess their student's language proficiency.

For example, Pearson's Versant suite of tests has been providing automated language assessment for almost 25 years. More than 350 million tests have been administered since its launch in 1996. The same technology is used in Pearson's Benchmark tests.

So what makes automated scoring systems reliable?

A large dataset of test answers and scores is used to train artificial intelligence machine learning technology to score English language tests in the same way as human marks. So we are not replacing human judgment; we are simply teaching computers to replicate it.

Of course, computers are much more efficient than humans. They are not averse to monotonous tasks and do not make mistakes (the standard error of the marking of an AI-scored test is lower than that of a human-scored test). So we can get unbiased, accurate and consistent scores.

The main advantages of automated scoring are speed, reliability, flexibility and uncertainty.

The main advantage of computers over humans is that they can process complex data quickly. Digital scoring often provides instant feedback. We can get accurate and reliable results in a matter of minutes. And this is not only for multiple-choice answers, but also for complex answers.

The advantage for teachers and institutions is that they can test hundreds, thousands, or tens of thousands of students at the same time and get instant scores. The faster you can collect scores, the faster you can make decisions about placement and student language levels, or assess student strengths and weaknesses and make adjustments to learning that lead to improvement and progress.

LITERATURE ANALYSIS AND RESEARCH METHODOLOGY

"Automated Essay Scoring: A Cross-disciplinary Perspective" Shermis, M.D., & Burstein, J. (2013) This book provides an excellent overview of automated essay scoring systems. The authors provide an in-depth analysis of the technologies, methodologies, and practices of AI-based written assessment. The book discusses the strengths and weaknesses of the systems, as well as a comparison of AI systems with human assessors. This is an important resource for English proficiency tests, as the writing section is often scored by AI.

"Artificial Intelligence in Education: Promises and Implications for Teaching and Learning" – Holmes, W., Bialik, M., & Fadel, C. (2019) This book examines the broad impact of AI in education. In particular, it provides detailed information on the role of AI in assessment systems. The authors focus on the issues of transparency, objectivity, and fairness in AI-based assessment. In the context of English language proficiency tests, an important literature for understanding the potential and potential risks of AI.

"Automated Scoring of Speaking Items: Prospects and Challenges" Xi, X. (2010) The article is specifically about the role and complexities of automated technologies in the assessment of spoken language. The difficulties and achievements of using AI to assess many factors such as pronunciation, intonation and logical expression in the assessment of the speaking section are shown. This article is very relevant in the analysis of the speaking section of English language tests.

"The Validity of Scores from Automated Scoring of Essays: Issues and Evidence" Williamson, D. M., Bejar, I. I., & Mislevy, R. J. (2006) The article examines the issues of reliability and validity of the results of automated essay scoring. The errors of the scoring systems, the bias in training the models and the interpretation of their results are discussed in detail. An essential resource for those concerned with test reliability and fairness.

"Exploring the Use of Artificial Intelligence in Language Assessment: A Review of Recent Developments" Zou, D., & Li, P. (2022) This article, based on the latest developments, reviews new trends and advanced methods in AI-based language assessment technologies. In particular, the application of multimodal training and assessment, machine learning, and natural language processing techniques to language tests is analyzed. Relevant literature for understanding the near future directions.[1; 62, 68-b]

These literatures show that:

AI increases the speed and objectivity of language test scoring.

Human control is still necessary in scoring, especially for tasks that require a high level of thinking and creative expression.

Technological developments offer great opportunities to make assessment systems more fair and transparent, but there is also the risk of bias and technical errors.

The rise of AI-based assessment systems requires new competencies in education and careers, meaning technological literacy is also becoming increasingly important.

ANALYSIS AND RESULTS

Determining English language proficiency is becoming increasingly important in today's era of globalization and digital transformation. Test systems (e.g. TOEFL, IELTS, Duolingo English Test) are widely used not only for academic and professional purposes, but also in immigration and employment processes. In recent years, there has been an increasing trend of using artificial intelligence (AI) and automated scoring technologies in these test systems. How are these technologies changing the process and what are their advantages and limitations?

AI-based systems allow for automatic assessment of test tasks. They are widely used in the following elements:

Reading and listening: Automatic identification of correct answers.

Writing: Semantic analysis of essays and short answers, assessment of grammatical accuracy and style.

Speaking: Assessment of pronunciation, fluency, grammatical correctness and logical expression.

AI systems combine machine learning, natural language processing (NLP), and big data analytics to provide results that are close to human-like.

1. Speed and efficiency: AI scoring provides results within seconds, which is much faster than traditional manual scoring.

2. Objectivity: Reduces the influence of subjective human opinions. The same criteria are used for each test taker.

3. Flexibility: Tests can be administered online and results are provided in real time.

4. Broad coverage: Allows you to take the test from anywhere in the world.

1. Underestimating complexity: AI may sometimes not fully understand contextual or creatively written answers.

2. Risk of unfairness: There is a possibility that programming errors or poorly trained models will give biased results to certain groups.

3. Distrust of artificial assessment: Some users still rely more on the impartiality and experience of human assessors.

4. Technical issues: Internet addiction, equipment quality, and technical failures can affect the results.[3; 345,358-b]

Duolingo English Test: This test is scored entirely by AI, and the result is ready in a few hours instead of several days.

TOEFL iBT: The writing and speaking sections are initially scored by AI, and then human assessors make the final decision.

IELTS Indicator: During the pandemic, an online version of IELTS was developed and integrated with AI-based analysis.

The development of AI technologies is expected to provide deeper and fairer assessment.

Multimodal assessment (combined analysis of text, audio, and video signals) can make test results more accurate and reliable.

Hybrid assessment systems that combine human and AI assessors are becoming more popular.

The next biggest benefit of digital assessments is flexible delivery models. As online learning becomes more popular, this is becoming increasingly important.

Availability has become a key issue: How can your institution provide access to assessments for your students if you can't deliver the test in a classroom setting?

For example, Versant, our web-based test, can be delivered online or offline, on-site or off-site. Test takers will need a computer and a headset with a microphone. They can take the test anywhere, any time of day, any day of the week, making it highly flexible to fit anyone's schedule or situation.

Impartiality is another key benefit of AI-based scoring. The AI engine used to score digital proficiency tests is completely free of bias. It doesn't get tired, it doesn't have good and bad days like human characters. And it doesn't have a personality.

While some human markers are more generous and others are stricter, AI is always equally fair. This means that automated scoring provides consistent, standardized scores regardless of who is taking the test.

If you are testing students from all over the world, they will be assessed solely on their English proficiency, in a completely objective manner.

Additional advantages of automated assessment are security and cost.

Digital assessments are more difficult to monitor than in-person tests, so security is a legitimate concern. One way to address this is through remote monitoring.

Remote monitoring adds an extra layer of security so test administrators can be sure that students taking the test from home are not cheating.

For example, our software captures video of test takers and an AI detection system automatically identifies suspicious test taker behavior. Test administrators can access the video at any time for audit and review and easily find suspicious segments highlighted by the AI.

Here are some examples of suspicious behavior that our system can detect:

Image monitoring:

Another face or multiple faces appear in the frame

Camera blocked

Browser monitoring:

Moving away from the test window or switching tabs repeatedly

Video monitoring:

Test taker moving out of camera view

Multiple people in camera view

Repeatedly looking away from the camera. [4; 182,190-b]

CONCLUSION

In short, AI and automated scoring are revolutionizing English language testing. They speed up the assessment process, increase objectivity, and create many new opportunities. However, these systems need to be carefully developed to keep them fair, transparent, and reliable, and not completely eliminate the human element.

Finally, the cost of automated English language certification is beneficial. Indeed, automated scoring can be a cost-effective way to monitor tests because it saves time and resources.

Pearson's English proficiency assessment is very scalable and does not require additional time from scorers, regardless of how many test takers you have.

In addition, there is no need to spend time and money on purchasing study markers or equipment.

AI helps you lead the way with efficient, convenient, fair, and cost-effective English test marking/administration. Over time, it should evolve, become more advanced, and provide even more support in the world of English language learning and assessment.

Classroom management: AI can help track student progress, identify at-risk students, and recommend intervention strategies before students fall behind.

Another important application of AI in ESL is real-time language proficiency testing. AI-powered platforms are used for a wide range of standardized tests as well as everyday classroom assessments. Some examples include:

Automated speaking tests: Platforms like Duolingo English Test or Pearson's PTE Academic use AI to assess students' speaking skills through instant feedback on pronunciation, fluency, and coherence.

Reading Comprehension Assessment: AI tools can analyze how well students understand written texts by evaluating their responses to comprehension questions, summarizing exercises, or filling in the blanks (fill-in-the-blank exercises).

Writing Assessment: AI is also used to grade written assignments in real time, assessing not only grammar but also content, coherence, and writing style.

Such real-time assessment offers students greater flexibility and accessibility, allowing them to take tests from anywhere and get instant results.

REFERENCES:

- Alderson, J. C. Assessing reading. Cambridge: Cambridge University Press. 2000.
- Alamri, H.A., Watson, S. & Watson, W. Learning Technology Models that Support Personalization within Blended Learning Environments in Higher Education. TechTrends 2021. 65, 62 78.
- Ashwathy Ashokan, Christian Haas, Fairness metrics and bias mitigation strategies for rating predictions, Information Processing & Management, Volume 58, Issue 5, 2021, 345-358
- Boni, M. The ethical dimension of human artificial intelligence collaboration. European View, 2021. 20(2), 182.190
- Burr Settles, Geoffrey T. aFlair, Masato Hagiwara; Machine Learning Driven Language Assessment. Transactions of the Association for Computational Linguistics 2020; 8 247,263
- Chao Han Translation quality assessment: a critical methodological review, The Translator, 2020. 26:3, 257 273,
- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. Application and theory gaps during the rise of Artificial Intelligence in Education. Computers 2020.
- Chalhoub-Deville, Micheline & Craig Deville. Computer adaptive testing in second language contexts. Annual Review of Applied Linguistics 1999. 273-299.