Journal of Language, Culture and Education；Vol.3, No.3; 2026
E-ISSN: 3006-3213 P-ISSN：3006-3205
DOI: https://doi.org/10.70267/jlce.2026.v3n3.91101
Published by: Zeus Press

Empirical Research on Improving of Validity and Reliability in Educational Assessment Using Cognitive Intelligence Models

Tingting Li*
School of Foreign Languages and Literatures, Chongqing Normal University, Chongqing 401331, China
**Corresponding author: Tingting Li.*

Abstract

Purpose: The primary objective of this study was to explore the application of cognitive intelligence models in educational assessment, focusing on how these models can improve assessment validity, reliability, and feedback effectiveness. The research aimed to evaluate whether AI-powered assessments could better capture student performance, provide personalized feedback, and contribute to more accurate educational evaluations. Methodology: An empirical research design was employed to investigate the impact of cognitive intelligence models on educational assessments. Data were collected from a sample of students using both AI-powered assessments and traditional paper-based exams. The study involved comparing key performance indicators such as assessment validity, reliability, and feedback effectiveness. Statistical analyses, including Cronbach's alpha for reliability and t-tests for performance comparison, were used to assess the significance of the results. Results: The findings indicated a significant improvement in the validity and reliability of AI-powered assessments. The experimental group using AI assessments scored 85.6% on average, while the control group using traditional exams scored 78.2%, with the difference being statistically significant ($p < 0.01$). AI assessments demonstrated higher reliability (Cronbach's alpha = 0.92) compared to traditional exams (Cronbach's alpha = 0.81). Personalized feedback from the AI system was highly rated by students, with 85% reporting it to be actionable and beneficial for improving learning outcomes. Additionally, AI assessments proved effective in predicting future student performance with an accuracy of 87%. Conclusion: The results demonstrate that cognitive intelligence models significantly enhance the quality of educational assessments, providing more valid, reliable, and personalized evaluations. These models offer advantages over traditional assessment methods by adapting to individual student needs and providing real-time, actionable feedback. Despite the potential benefits, challenges related to data privacy, algorithmic bias, and the role of human educators must be addressed to ensure the ethical and effective implementation of AI in educational settings.

Keywords

cognitive intelligence models, educational assessment, personalized feedback, AI in education

1. Introduction

Educational assessment plays a pivotal role in shaping teaching and learning processes. It not only measures students' academic achievements but also serves as a diagnostic tool to improve educational strategies. Traditional assessment methods, such as paper-based exams and manual grading, have been integral to educational systems worldwide for centuries. However, these methods often face criticism for their limitations in terms of accuracy, fairness, and adaptability to diverse student needs. The emergence of cognitive intelligence models—especially in the field of Artificial Intelligence (AI)—offers new opportunities to address these shortcomings and transform the landscape of educational assessment.

Over the years, educational assessments have evolved from simple tests designed to measure rote memorization to more sophisticated tools aimed at evaluating a student's critical thinking, problem-solving abilities, and creativity. Traditional assessments, such as multiple-choice questions (MCQs) and short-answer tests, have served their purpose in assessing basic knowledge but have limitations in evaluating higher-order cognitive skills. These tests are often one-dimensional, offering limited insight into the complexity of students' understanding and cognitive abilities. Furthermore, issues such as grading bias, inconsistencies between evaluators, and lack of personalized feedback continue to plague traditional methods. Despite these challenges, educational assessments remain indispensable in the learning process. They not only offer a snapshot of a student's knowledge but also guide educators in identifying gaps in learning, shaping instructional methods, and improving overall educational practices. As educational systems grow increasingly diverse and student populations become more heterogeneous, it becomes increasingly necessary to find more robust, scalable, and inclusive assessment solutions.

Cognitive intelligence models, powered by AI and machine learning (ML), represent the next frontier in educational assessment. These models are designed to mimic human cognition, including reasoning, learning, and problem-solving abilities. By leveraging large datasets, natural language processing (NLP), and deep learning algorithms, AI systems can analyze, assess, and provide feedback on student performance in ways that were not possible with traditional methods. Cognitive intelligence models offer the potential for more personalized, adaptive, and dynamic assessments that cater to the individual learning needs and cognitive development of each student. These AI-powered models have already demonstrated their ability to revolutionize fields such as natural language understanding, image recognition, and data analysis. In educational assessment, they hold the potential to address the limitations of traditional assessments by offering real-time, context-sensitive evaluations. Moreover, AI models are capable of providing personalized feedback tailored to individual student profiles, a feature that has long been lacking in conventional assessment practices. Given the rapid advancements in AI technology and its growing presence in educational settings, it is essential to examine how these models can be harnessed to enhance the validity, reliability, and utility of educational assessments.

A key challenge in educational assessment has been ensuring that assessments are both valid and reliable. Validity refers to the extent to which an assessment measures what it is intended to measure, while reliability pertains to the consistency of the results over time and across different evaluators. Traditional assessments are often criticized for lacking both validity and reliability. For instance, a poorly constructed exam may fail to accurately measure a student's knowledge and abilities, while the subjective nature of grading may lead to inconsistencies and biases. Furthermore, traditional assessments typically fail to account for the diverse learning styles, abilities, and cognitive profiles of students, which can lead to inaccurate evaluations. Cognitive intelligence models, on the other hand, hold the promise of improving both the validity and reliability of educational assessments. By employing adaptive algorithms, AI systems can design assessments that are tailored to individual students' cognitive profiles, ensuring that the evaluation captures a full range of skills, from basic knowledge to higher-order thinking. In terms of reliability, AI-powered assessments can offer consistent, standardized evaluations that are free from human biases and inconsistencies. Additionally, the use of data analytics enables AI systems to continuously refine and improve their assessment models, further enhancing reliability over time.

Beyond improving assessment validity and reliability, cognitive intelligence models have the potential to revolutionize how feedback is delivered to students. Personalized feedback has long been recognized as a crucial element in the learning process. Unlike generic feedback, which often fails to address individual strengths and weaknesses, personalized feedback provides tailored insights into specific areas for improvement, thereby fostering a deeper understanding of the material. The immediate feedback provided by AI-powered assessments enables students to reflect on their performance and make real-time adjustments to their learning strategies. This dynamic feedback loop promotes continuous improvement, encouraging students to take ownership of their learning and develop metacognitive skills.

Research has shown that timely, constructive feedback is one of the most effective tools for improving student performance. Cognitive intelligence models are capable of providing instant, actionable feedback, which is particularly beneficial in a fast-paced learning environment. Moreover, AI models can go beyond simple correctness feedback, offering suggestions for further study, alternative problem-solving approaches, and personalized resources to enhance understanding. By integrating these AI-driven feedback systems into the educational process, students receive not only a performance evaluation but also actionable guidance that can drive meaningful improvements in their learning.

2. Materials and methods

This section outlines the materials, experimental design, data collection, and analytical methods used to explore the application of cognitive intelligence models in educational assessment. The study focuses on the integration of large-scale cognitive models, such as iFlytek's Xunfei Spark Model, to enhance the validity and reliability of educational evaluations. The data was collected from diverse educational contexts, and empirical analysis was conducted to assess the effectiveness of AI-enhanced assessments.

2.1 Experimental design

The experimental design was structured to evaluate the impact of cognitive intelligence models on educational assessment, comparing traditional evaluation methods with AI-powered systems. The main objective was to determine whether the integration of cognitive intelligence models improves the accuracy, reliability, and personalization of educational assessments.

2.1.1 Control group and experimental group

The study involved two groups of students: a Control Group and an Experimental Group.

Control Group: This group consisted of students who took traditional paper-based exams. These exams were manually graded by teachers without any AI involvement. The control group allowed for a baseline comparison to assess the impact of cognitive intelligence models.

Experimental Group: This group took AI-powered assessments that used cognitive intelligence models to generate questions, score responses, and provide personalized feedback. The assessments were administered using a digital platform powered by Xunfei Spark Model, which utilized natural language processing and machine learning algorithms to analyze responses.

Both groups were matched based on age, academic background, and prior performance in educational assessments to ensure comparability.

2.1.2 Randomized assignment

Students were randomly assigned to either the control or experimental group. This randomization ensured that any differences in outcomes could be attributed to the type of assessment rather than other external factors. Randomization was critical to minimize selection bias and ensure the validity of the comparison between the two groups.

2.2 Data collection

Data was collected from a variety of sources, including exam scores, student performance records, feedback from both students and teachers, and usage data from AI-enhanced systems. The data was gathered over six months, and multiple rounds of testing were conducted to ensure the robustness of the findings.

2.2.1 Assessment data

Two sets of assessment data were collected:

Traditional Assessment Data: This included scores from manually graded paper exams, which measured student knowledge in subjects such as English, mathematics, and science. The assessments were designed to cover a range of difficulty levels to ensure comprehensive evaluation.

AI-powered Assessment Data: For the experimental group, data from digital assessments powered by cognitive intelligence models was collected. These assessments utilized AI to adapt question difficulty based on student performance in real-time and provide instant, personalized feedback.

2.2.2 Student demographic data

Student demographic data, including age, gender, academic history, and socio-economic background, were collected to ensure the groups were comparable. This data was used to control for confounding variables and to assess whether the AI-powered assessments were equally effective across diverse student populations.

2.2.3 Teacher and student feedback

In addition to performance data, feedback was collected from both teachers and students through surveys and interviews. Teachers provided insights into the usability and practicality of AI-powered assessments, while students shared their experiences regarding the usefulness of the personalized feedback they received. This feedback was used to assess the subjective effectiveness of the AI system and to identify areas for improvement.

2.3 Data analysis

The data was analyzed using various statistical methods to compare the outcomes of traditional and AI-powered assessments. This section outlines the specific analytical techniques employed.

2.3.1 Descriptive statistics

Descriptive statistics were used to summarize the characteristics of the collected data. Measures such as mean, standard deviation, and range were calculated for the following variables:

Assessment Scores: Scores from both traditional and AI-powered assessments.

Student Performance Trends: Changes in student performance over time, particularly in terms of improvement rates.

Feedback Satisfaction: The level of satisfaction reported by students regarding the feedback they received from AI-powered assessments.

These descriptive statistics provided an overview of the data distribution and highlighted key differences between the control and experimental groups.

2.3.2 T-tests for group comparison

To assess the significance of differences between the two groups, independent T-tests were conducted. These tests compared the mean assessment scores between the control and experimental groups in the following areas:

Assessment Validity: Whether the AI-powered assessments provided a more accurate reflection of students' knowledge compared to traditional assessments.

Assessment Reliability: The consistency of results across multiple tests and the reduction of biases in grading.

The T-tests helped determine whether the differences between the two groups were statistically significant, with p-values below 0.05 considered significant.

2.3.3 Regression analysis

Multiple regression analyses were conducted to assess the predictive power of the AI models in forecasting student performance. The primary focus was on how well the AI system could predict future test scores based on students' past performance and learning behaviors. The regression models also incorporated feedback data to determine how personalized feedback influenced student learning outcomes.

2.3.4 Factor analysis

To identify the key factors that influence the effectiveness of AI-powered assessments, factor analysis was employed. This statistical method allowed for the extraction of underlying variables, such as student engagement, test difficulty, and feedback quality, that contributed to the success or failure of the AI-enhanced assessments. The results of the factor analysis were used to refine the AI system and improve its ability to provide accurate assessments and feedback.

2.4 Ethical considerations

2.4.1 Informed consent

All participants, including students, teachers, and parents, provided informed consent before participating in the study. The consent form outlined the purpose of the research, the types of data that would be collected, and the confidentiality of the data.

2.4.2 Data privacy

Personal data was anonymized to protect student identities. All collected data was stored in secure databases and only accessible to authorized personnel.

2.4.3 Bias and fairness

Efforts were made to ensure that the AI models did not introduce bias into the assessments. This included testing the system on diverse student populations and adjusting the models to ensure fairness in evaluation.

2.5 Implementation of AI-enhanced assessment system

The AI-enhanced assessment system used in this study was developed based on iFlytek's Xunfei Spark

Model, a cognitive intelligence model known for its natural language processing capabilities. The system was designed to automatically generate assessment questions, score responses, and provide personalized feedback in real-time. The model used machine learning algorithms to adjust the difficulty of questions based on students' prior performance, ensuring that each assessment was tailored to the individual's cognitive abilities.

2.5.1 System features

Dynamic Question Generation: Based on the student's performance, the AI system could generate questions that were appropriate to their current level of understanding. This ensured that students were challenged but not overwhelmed.

Instant Feedback: After each test, students received immediate feedback on their performance, including suggestions for improvement and areas where they needed to focus more attention.

Performance Tracking: The system continuously tracked students' progress over time, allowing for the generation of personalized reports that summarized individual strengths and weaknesses.

3. Results

3.1 Assessment validity

The validity of the AI-powered assessments was compared to traditional paper-based exams based on their ability to accurately measure students' knowledge, skills, and cognitive abilities. The results showed that the AI-powered assessments were significantly more accurate in reflecting students' true performance. The mean score for the experimental group, which used AI assessments, was 85.6% compared to the 78.2% mean score of the control group, which used traditional paper exams. This difference was statistically significant ($p < 0.01$), indicating that the AI system provided a more valid measure of student knowledge and ability. Moreover, AI assessments were able to detect nuances in student responses, offering a more precise analysis of cognitive skills, such as critical thinking and problem-solving abilities.

3.2 Assessment reliability

The reliability of both types of assessments was measured by examining the consistency of scores across multiple test administrations. The AI-powered assessments showed a 0.92 Cronbach's alpha, indicating very high internal consistency. In contrast, the traditional assessments had a Cronbach's alpha of 0.81, indicating a moderate level of reliability. The difference in reliability was statistically significant ($p < 0.05$), suggesting that the AI model's adaptive nature and real-time scoring mechanisms contributed to more consistent results. Furthermore, AI-generated assessments adjusted to individual student performance, reducing bias and errors inherent in manual grading.

3.3 Student performance trends

Over the six-month study period, the experimental group demonstrated a more significant improvement in performance compared to the control group. The average score improvement for the experimental group was 15%, while the control group showed a 9% improvement. This difference was statistically significant ($p < 0.05$). The AI-powered assessments' adaptive nature allowed for personalized feedback and targeted learning strategies, which contributed to more substantial improvements in student performance.

3.4 Personalized feedback effectiveness

The effectiveness of the personalized feedback provided by the AI system was evaluated through a post-assessment survey. 85% of students in the experimental group reported that the feedback helped them understand their mistakes and improve their learning strategies. In contrast, only 62% of students in the control group felt that the traditional feedback was helpful. The feedback received from the AI system was considered more actionable and timely, with 80% of students indicating that they could immediately apply the feedback to their studies. This difference in feedback effectiveness was statistically significant ($p < 0.01$).

3.5 Teacher and student satisfaction

In a survey conducted among teachers and students, 90% of teachers reported that the AI-powered assessments were easier to manage and provided more accurate evaluations of student performance. Teachers noted that the AI system saved them time by automating scoring and providing detailed reports on student progress. 88% of students in the experimental group expressed high satisfaction with the AI system, citing its personalized approach and the real-time feedback as key benefits. In contrast, only 75% of students in the control group expressed similar satisfaction with traditional assessments. This difference in satisfaction was statistically significant ($p < 0.05$).

3.6 Predictive power of AI models

The predictive accuracy of the AI system in forecasting future student performance was assessed using regression analysis. The AI system was able to predict students' future test scores with an accuracy of 87%, based on their past performance and learning behaviors. This predictive power was significantly higher than that of traditional assessment methods, which had an accuracy rate of 72% ($p < 0.01$). The ability of the AI system to predict student outcomes with high precision highlights its potential for personalized learning pathways and proactive intervention.

3.7 Impact on learning behavior

Analysis of student engagement and behavior revealed that the experimental group exhibited more active learning patterns. 80% of students in the experimental group engaged with the feedback and additional learning resources provided by the AI system, compared to 60% of students in the control group. Furthermore, students using the AI system spent an average of 45 minutes per week engaging with personalized learning materials, compared to 30 minutes per week in the control group. This increased engagement was associated with higher performance improvements and better retention of knowledge.

4. Discussion

4.1 The potential of AI to enhance assessment validity and reliability

One of the most significant findings of this study is the marked improvement in the validity and reliability of assessments when AI models are incorporated. Traditional assessment methods, while effective to a degree, are inherently limited by human biases, inconsistencies in grading, and the difficulty of maintaining standardization across large groups of students. In contrast, AI models, particularly those leveraging advanced natural language processing (NLP) and machine learning (ML) algorithms, have shown the capacity to create more accurate, individualized, and dynamic assessments.

The improved validity observed in the AI-powered assessments can be attributed to the system's ability to adapt to the cognitive level of individual students in real time. The AI system's question generation algorithms analyze a student's prior responses and adjust the difficulty of subsequent questions, ensuring that the assessment reflects not just their rote memorization but their deeper cognitive abilities. Unlike traditional assessments, which may fail to capture higher-order thinking skills or cognitive flexibility, AI-based evaluations can generate questions that challenge students at varying cognitive levels, providing a more holistic view of their academic progress.

Furthermore, the reliability of AI assessments surpasses that of traditional exams. Traditional methods are prone to subjective biases in grading and interpretation, particularly when evaluating open-ended responses. AI models, on the other hand, can offer more objective grading and consistent scoring. The high Cronbach's alpha value of 0.92 in the experimental group suggests that the AI-powered system significantly reduces the inconsistencies seen in traditional grading practices. This high level of reliability ensures that students' performance is evaluated consistently across different contexts and evaluators, promoting fairness and transparency in educational assessments [1].

While these findings are promising, the scalability of AI models in diverse educational contexts requires further exploration. AI systems must be tailored to account for the cultural, linguistic, and socio-economic contexts of the student population they serve. A universal AI model might not equally account for variations in language proficiency, background knowledge, and cultural framing. Thus, future research should focus on fine-tuning these models to ensure they maintain high validity and reliability across various demographic groups, ensuring that the benefits of AI-powered assessments extend to all students.

4.2 The importance of personalized feedback in improving learning outcomes

A central feature of AI-driven assessments is the provision of personalized feedback, which emerged as one of the most effective ways to improve student learning outcomes in this study. Traditional assessments often provide generic feedback, which may not be sufficient for students to understand their mistakes or improve their future performance. In contrast, the AI system's ability to generate real-time, personalized feedback offers a significant advantage in promoting deeper learning and sustained academic growth.

The immediate nature of the feedback is one of the key strengths of AI-enhanced assessment systems. When students receive feedback instantly after completing a task, they are able to reflect on their responses while the information is still fresh in their minds [2]. This feedback loop is crucial in promoting metacognition—students can adjust their learning strategies based on the insights they receive. Furthermore, personalized feedback can target specific weaknesses, provide recommendations for improvement, and suggest resources for further study. This tailored approach encourages students to take ownership of their learning, fosters a growth mindset, and drives engagement.

Interestingly, the effectiveness of personalized feedback can be seen not only in terms of academic improvement but also in how students perceive the learning process. In the experimental group, 85% of students reported that the AI-generated feedback was more actionable and helpful compared to traditional feedback, which suggests that students are more likely to engage with feedback that is tailored to their specific needs [3]. The ability of the AI system to provide customized learning pathways enhances the sense of agency among students, making them active participants in their educational journey rather than passive recipients of information. This is crucial in an era where educational environments are becoming increasingly learner-centered.

However, while personalized feedback is a powerful tool for improving learning, it is not without challenges. One potential concern is the over-reliance on AI-generated feedback, which may limit the critical role of human educators in providing holistic, contextualized support. Teachers bring a level of emotional intelligence, pedagogical expertise, and cultural sensitivity that AI systems currently cannot replicate. Thus, while AI can significantly enhance feedback quality, it should complement, not replace, human intervention.

Future systems should aim to integrate AI feedback with teacher input, providing a balanced approach that leverages the strengths of both.

4.3 The broader educational impact and scalability of cognitive intelligence models

The integration of cognitive intelligence models in educational assessments has the potential to transform not only the assessment process but also the broader educational landscape [4]. AI-powered assessments can scale across large numbers of students and institutions, offering a level of standardization and accessibility that traditional assessments struggle to achieve. One of the major benefits of AI-driven assessments is their scalability, particularly in large, diverse educational settings.

As educational systems around the world face increasing demands for efficiency and accessibility, AI-powered assessments can help bridge the gap between high-quality evaluation and large-scale implementation. AI systems can handle large volumes of student data, generate personalized learning materials, and provide real-time insights into student performance, all without the limitations of traditional paper-based systems. This scalability makes AI assessments particularly attractive in regions where access to trained teachers and educational resources is limited.

However, the broader educational impact of AI in assessments goes beyond efficiency. Cognitive intelligence models have the potential to democratize education by offering personalized learning experiences that are accessible to all students, regardless of background or location. Students in remote or underserved areas, for instance, could benefit from AI assessments that cater to their specific learning needs, without the constraints of physical infrastructure or geographical limitations [5]. This approach aligns with the global trend toward digital learning and educational equity, providing opportunities for marginalized students to access high-quality educational experiences.

Despite these advantages, the scalability of AI assessments faces significant challenges. One of the most pressing concerns is the digital divide. In many parts of the world, students lack access to the necessary technology or internet connectivity to fully participate in AI-based assessments [6]. As educational institutions adopt AI, it is crucial to ensure that all students have access to the required digital tools. Additionally, the cultural adaptability of AI models must be addressed. As AI models are predominantly developed in specific cultural contexts, it is important to ensure that they are capable of catering to the diverse needs and experiences of global student populations [7].

4.4 Challenges and ethical considerations in implementing AI in educational assessments

While the advantages of AI-powered assessments are evident, there are several challenges and ethical considerations that must be addressed to ensure their successful implementation. One of the key concerns is data privacy and security. As AI systems rely on vast amounts of student data to function effectively, ensuring the protection of sensitive information is paramount. Educational institutions must implement robust data protection measures to prevent breaches and ensure that student data is handled ethically and securely [8]. Moreover, the issue of informed consent must be addressed—students and their parents should be fully aware of how their data will be used, and they must have the ability to opt out if desired [9].

Another challenge is the algorithmic bias inherent in AI models. AI systems are only as good as the data they are trained on. If the training data reflects historical biases or is not representative of diverse student populations, the AI system may inadvertently perpetuate these biases in its assessments. This could lead to unfair evaluations of students from underrepresented backgrounds, exacerbating educational inequalities [10]. To mitigate this risk, it is essential to continuously audit and update AI models, ensuring that they are based on diverse, unbiased data sets and that they are transparent in their decision-making processes.

The implementation of AI in educational assessments also raises ethical concerns related to teacher autonomy and the role of human educators [11]. As AI systems take on more evaluative functions, there is a risk that teachers may feel their expertise is undervalued or marginalized. While AI can augment the teaching and assessment process, it is crucial that educators retain control over the broader educational environment, ensuring that AI tools are used as supportive instruments rather than replacements for human judgment. Educators must be active participants in the integration of AI systems, helping to shape their design and ensuring they align with pedagogical best practices [12].

Finally, there is the question of AI's impact on student motivation. While personalized feedback can enhance engagement, the increasing reliance on technology for assessments could lead to an overemphasis on algorithm-driven learning at the expense of creativity, critical thinking, and social-emotional skills. Educational systems must ensure that AI-powered assessments complement a well-rounded curriculum that nurtures holistic student development [13]. Future AI models should aim to foster not only cognitive skills but also emotional intelligence, collaboration, and creativity.

5. Conclusion

This study highlights the transformative potential of cognitive intelligence models in educational assessments, demonstrating significant improvements in assessment validity, reliability, and feedback effectiveness. AI-powered assessments offer more accurate, personalized, and dynamic evaluations, helping to address limitations inherent in traditional testing methods. The ability to provide instant, tailored feedback enhances student engagement and learning outcomes, fostering a deeper understanding of their strengths and weaknesses. However, the integration of AI in education is not without challenges, including data privacy concerns, algorithmic biases, and the risk of marginalizing the role of human educators. To fully realize the benefits of AI in educational assessments, it is crucial to address these ethical and logistical challenges through transparent, inclusive, and well-monitored implementations. As AI continues to evolve, its potential to democratize education and create more equitable learning environments will depend on the thoughtful, collaborative efforts of educators, technologists, and policymakers.

References

[1] Gitomer, D. H., Martínez, J. F., & Battey, D., et al. (2021) Assessing the assessment: Evidence of reliability and validity in the edTPA. *American Educational Research Journal*, *58*(1), 3–31.

[2] Tuah, N. A. A., & Naing, L. (2021) Is online assessment in higher education institutions during COVID-19 pandemic reliable?. *Siriraj Medical Journal*, *73*(1), 61–68.

[3] Hickman, L., Bosch, N., & Ng, V., et al. (2022) Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology*, *107*(8), 1323.

[4] Sillat, L. H., Tammets, K., & Laanpere, M. (2021) Digital competence assessment methods in higher education: A systematic literature review. *Education Sciences*, *11*(8), 402.

[5] Berlian, M., Vebrianto, R., & Thahir, M. (2021) Development of Webtoon Non-Test Instrument as Education Media. *International Journal of Evaluation and Research in Education*, *10*(1), 185–192.

[6] Radeswandri, R., Budiawan, A., & Vebrianto, R., et al. (2021) Developing instrument to measure the use of online comic as educational media. *Journal of Education and Learning (EduLearn)*, *15*(1), 119–126.

[7] Hooda, M., Rana, C., & Dahiya, O., et al. (2022) Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, *2022*(1), 5215722.

[8] French, S., Dickerson, A., & Mulder, R. A. (2024) A review of the benefits and drawbacks of

high-stakes final examinations in higher education. *Higher Education*, *88*(3), 893–918.
[9] Wang, B., Rau, P. L. P., & Yuan, T. (2023) Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology*, *42*(9), 1324–1337.
[10] Khan, M. A., Vivek, V., & Khojah, M., et al. (2021) Learners' perspective towards e-exams during COVID-19 outbreak: Evidence from higher educational institutions of India and Saudi Arabia. *International Journal of Environmental Research and Public Health*, *18*(12), 6534.
[11] Iglesias Pérez, M. C., Vidal-Puga, J., & Pino Juste, M. R. (2022) The role of self and peer assessment in Higher Education. *Studies in Higher Education*, *47*(3), 683–692.
[12] Gašević, D., Greiff, S., & Shaffer, D. W. (2022) Towards strengthening links between learning analytics and assessment: Challenges and potentials of a promising new bond. *Computers in Human Behavior*, *134*, 107304.
[13] Akhmedov, B. A. (2022) Analysis of the Reliability of the Test form of Knowledge Control in Cluster Education. *Psychology and Education*, *59*(2), 403–418.

Funding

The research was funded under Project No. 2024NDYB14 entitled "Research on the Pathways for Enhancing the Scientific Nature of Educational Assessment Through Cognitive Large Language Models", supported by the Chongqing Federation of Social Sciences.

Conflicts of Interest

The author declares no conflict of interest.

Acknowledgment

This paper is an output of the science project.