



A Systematic Review of AI-Driven Educational Assessment in STEM Education

Fan Ouyang¹ · Tuan Anh Dinh¹ · Weiqi Xu¹

Accepted: 20 October 2023 / Published online: 9 November 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Artificial intelligence (AI), as an emerging technology, has been widely used in STEM education to promote the educational assessment. Although AI-driven educational assessment has the potential to assess students' learning automatically and reduce the workload of instructors, there is still a lack of review works to holistically examine the field of AI-driven educational assessment, especially in the STEM education context. To gain an overview of the application of AI-driven educational assessment in STEM education, this research conducted a systematic review based on 17 empirical research published from 2011 January to 2023 April. Specifically, this review examined the functions, algorithms, and effects of AI applications in STEM educational assessment. The results clarified three main functions of AI-driven educational assessment, namely academic performance assessment, learning status assessment, and instructional quality assessment. Moreover, the systematic review found that both traditional algorithms (e.g., natural language processing, machine learning) and advanced algorithms (e.g., deep learning, neural fuzzy systems) were applied in STEM educational assessment. Furthermore, the educational and technological effects of applying AI-driven educational assessment in STEM education were revealed. Based on the results, this research proposed educational and technological implications to guide the future practice and research of AI-driven educational assessment in STEM education.

Keywords Artificial intelligence · Artificial intelligence in education · Educational assessment · STEM education

✉ Fan Ouyang
fanouyang@zju.edu.cn

¹ College of Education, Zhejiang University, #866, Yuhangtang Rd., Hangzhou 310058, Zhejiang, China

Artificial Intelligence in STEM Education

STEM education, as an interdisciplinary learning approach in education, focuses on the integration of science, technology, engineering, and mathematics, in order to help students develop interdisciplinary knowledge through solving real-world problems, enhance their higher-order thinking and collaboration skills, and transform education from the teacher-directed lectures to active learner-centered learning (Henderson et al., 2011; Kennedy & Odell, 2014). However, with the development of STEM education, it usually faces challenges during the instructional and learning processes, such as tracking students' learning processes and attitudes (Vennix et al., 2018), designing authentic, productive STEM problems (Theobald et al. 2020), and assessing their performances (Jiao et al., 2022). To address these issues, emerging technologies, particularly artificial intelligence (AI), have been used in STEM education to decrease instructors' workload and promote students' personalized and adaptive learning (Chen, et al., 2020; Hwang et al., 2020; Jiao et al., 2022; Ouyang & Jiao, 2021). Specifically, the application of AI technologies has great potential to bring opportunities for innovations in STEM education, such as providing recommendations and feedback to students (Debus & Lawley, 2016), identifying at-risk students (Holstein et al., 2018), providing learning materials based on students' needs (Chen et al., 2020), and assessing their learning automatically (Wang et al., 2011; Zampiroli et al., 2021). Tseng et al. (2023) utilized machine learning and physiological signals to assess students' personality traits and aided students in improving their academic performance. Overall, AI technologies have been widely applied in STEM education to enhance the efficiency of instruction, empower students' learning, and transform the educational system.

Relevant Work

Educational Assessment in STEM Education

Educational assessment, as a process of gathering and analyzing information about students' learning performance, is usually applied in STEM education to help instructors gain understanding of students' learning status and make instructional decisions (Conley, 2015). Specifically, educational assessment mainly includes formative and summative assessment in STEM education. First, in formative assessment, the assessor (e.g., instructor, educational researcher) takes an active approach to provide ongoing feedback on students' performance during the learning process, which helps students understand their learning status to further modify their learning behaviors (Hendriks et al., 2019). Second, summative assessment is conducted at the end of the instructional and learning phase, which is primarily used for the final scoring and accreditation, usually without providing further feedback or adjustment opportunities for students (Lin & Lai, 2014). In

STEM education, formative and summative assessments are usually used together to track and evaluate students' learning behaviors and performance during the instructional and learning process (Chandio et al. 2016). However, some challenges and problems exist during the educational assessment process in STEM education. For example, during the formative assessment in STEM education, one of the challenges is how to provide adequate and timely feedback to students about their learning performance (Tempelaar et al. 2015). In addition, the substantial and repetitive tasks during summative assessment also bring huge workload to instructors in STEM education, which might reduce the time they spend on the personalized student-centered instruction. Therefore, how to increase the effects of educational assessment has become a critical problem to promote the development of STEM education.

AI-Driven Educational Assessment in STEM Education

The emergence and application of AI techniques brings great opportunities to reshape educational assessment in STEM education (Ouyang et al., 2022; Wang et al., 2011; Zampirolli et al., 2021). AI-driven educational assessment highlights using automated AI algorithms and models to assess or evaluate the instructional and learning process, in order to improve the assessment efficiency and promote the quality of STEM education (Gobert et al., 2013). First, in formative assessment, previous researchers have started to use AI techniques to support the assessment process through automated data collection and analysis. For example, Saito and Watanobe (2020) proposed a learning path recommendation system designed with natural language processing (NLP) to assess students' programming learning performance. Xing et al. (2021) used Bayesian networks to automatically assess students' engineering learning performance during the STEM learning process. Through using AI techniques, instructors can track and understand students' learning status, and the issue of delayed feedback in formative assessment can be solved. Second, in summative assessment, AI technologies can also assist instructors to automatically assess students' STEM learning performance. In addition, through analyzing and synthesizing large amounts of data, AI algorithms can provide insights and reveal students' learning patterns that may not be immediately obvious to instructors (Ekolu, 2021). For example, Erickson et al. (2020) applied an NLP-enabled automated assessment system in a mathematics curriculum to assess students' learning performance. Yang (2023) presented an approach for assessing undergraduate students' digital literacy through using natural language processing and other AI technologies. Bertolini et al. (2023) argued that Bayesian methods offered a practical and interpretable way to evaluate students' learning performance, which was a promising AI model for science education research and assessment. Overall, the application of AI technologies has the potential to empower the formative and summative assessment process, with an ultimate goal to enhance the quality of STEM education.

Previous Reviews and the Purpose of Current Study

Recent literature review works have started to focus on the field of AI-driven educational assessment. Through database searching, we located two systematic reviews about the application of AI in educational assessment (i.e., Gardner et al., 2021; Zehner & Hahnel, 2023). Specifically, Gardner et al. (2021) stated the progress AI had made in educational assessment and further hoped that AI can mimic human judgment and make breakthroughs in formative assessment. Zehner and Hahnel (2023) discussed the application in using AI technologies in educational assessment, such as log data analysis, natural language processing, machine learning, and other methods in modern technology-based assessment approaches. Although these two reviews contributed to the understandings of AI-driven educational assessment, there is a lack of research focusing on the application of AI techniques in educational assessment under the STEM educational context. Therefore, the current study aimed to conduct a systematic review, in order to clarify the functions, algorithms, and effects of AI-driven educational assessment in STEM education. To be specific, this systematic review focused on the following three research questions:

RQ1: What are the functions of AI applications in STEM educational assessment?

RQ2: What AI algorithms are used to achieve those functions in STEM educational assessment?

RQ3: What are the effects of AI applications in STEM educational assessment?

Methodology

We aimed to conduct a systematic review to holistically examine the application of AI-driven educational assessment in STEM education. The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) principles (Moher et al., 2009) were applied in this review work.

Database Search

We searched and filtered the articles in the main publisher databases: Web of Science, EBSCO, ACM, Wiley, Scopus, Taylor & Francis, IEEE. In addition, to locate the latest research about AI-driven educational assessment, we selected and searched the articles in the past decade, which were published from January 2011 to April 2023. The educational empirical research articles about AI-driven educational assessment from January 2011 to April 2023 were located. After database searching, a snowballing approach was utilized to identify the articles that were not captured by the initial search strings (Wohlin, 2014). Google Scholar was utilized in snowballing to search for the other articles.

Search Term

We adopt a structured search strategy to use different search methods according to the rules of different databases. Specifically, four types of keywords were selected as search terms, including keywords related to AI (i.e., “artificial intelligence” OR “AI” OR “AIED” OR “machine learning” OR “intelligent tutoring system” OR “expert system” OR “recommended system” OR “recommendation system” OR “feedback system” OR “personalized learning” OR “adaptive learning” OR “prediction system” OR “student model” OR “learner model” OR “data mining” OR “learning analytics” OR “prediction model” OR “automated evaluation” OR “automated assessment” OR “robot” OR “virtual agent” OR “algorithm” OR “intelligent tutoring system” OR “expert system” OR “expert system” OR “prediction model” OR “decision tree” OR “machine learning” OR “neural network” OR “deep learning” OR “k-means” OR “random forest” OR “support vector machines” OR “logistic regression” OR “fuzzy-logic” OR “Bayesian network” OR “latent Dirichlet allocation” OR “natural language processing” OR “genetic algorithm” OR “genetic programming”), keywords related to education (i.e., “education” OR “learning” OR “course” OR “class” OR “teaching”), keywords related to STEM (i.e., “STEM” OR “science” OR “technology” OR “math” OR “physics” OR “chemistry” OR “biology” OR “geography” OR “engineering” OR “programming” OR “lab”), and keywords related to assessment (i.e., “assessment” OR “evaluation”). It is worth noting that, based on the definition of STEM education, STEM education in this research includes science, technology, engineering, math, and cross-discipline (e.g., math and technology).

Screening Criteria

The inclusion and exclusion criteria were proposed to locate the articles about AI-driven educational assessment in STEM education. Based on the research objectives, a series of inclusion and exclusion criteria were proposed (see Table 1).

Table 1 The inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
1. The research topic must be relative to AI-driven educational assessment in STEM education	1. Research published as conference proceeding, book chapter, magazine, news is excluded
2. Research must be empirical research, which reports the application effects of AI-driven educational assessment in STEM learning process	2. Uncompleted research is excluded
3. Research must be published in peer-reviewed journals	3. Research that only reports AI application designs without empirical results is excluded
4. Research must be written in English	4. Empirical research that merely uses self-reported data, such as interview data is excluded
5. The full-text is available	

The Searching and Screening Process

Five steps were included in the searching and screening process: (1) removing duplicate articles, (2) reading the titles and abstracts and excluding articles based on the inclusion and exclusion criteria, (3) re-evaluating full texts and excluding articles based on the inclusion and exclusion criteria, (4) utilizing the snowballing approach to further locate articles, and (5) extracting and analyzing data from the final included articles (see Fig. 1). During the searching and screening process, all articles were imported into the Zotero software.

First, during the database searching and snowballing, we identified 464 articles based on the searching terms. After removing the duplicate articles, 453 articles were left. Second, through the titles and abstracts screening, the articles that did not meet the inclusion criteria were removed, leaving a total of 57 articles. The first author independently reviewed approximately 30% of the articles to ensure reliability. Then, the articles were read by the second author to crosscheck the reliability. The initial interrater agreement was 90% and reached 100% agreement after the discussion between the two authors. Subsequently, both authors reviewed the full texts of the articles to verify their compliance with all inclusion criteria for the review. Finally, 17 articles that met the criteria were included in the systematic review.

Data Analysis

We utilized the bibliometric analysis approach (Neuendorf & Kumar, 2015) to analyze the 17 included articles. To extract the data relevant to the research questions, we employed a qualitative content analysis approach to analyze the articles (Zupic & Čater, 2015). Specifically, two researchers discussed and verified the categorization of the reviewed articles (Graneheim & Lundman, 2004). After clarifying the

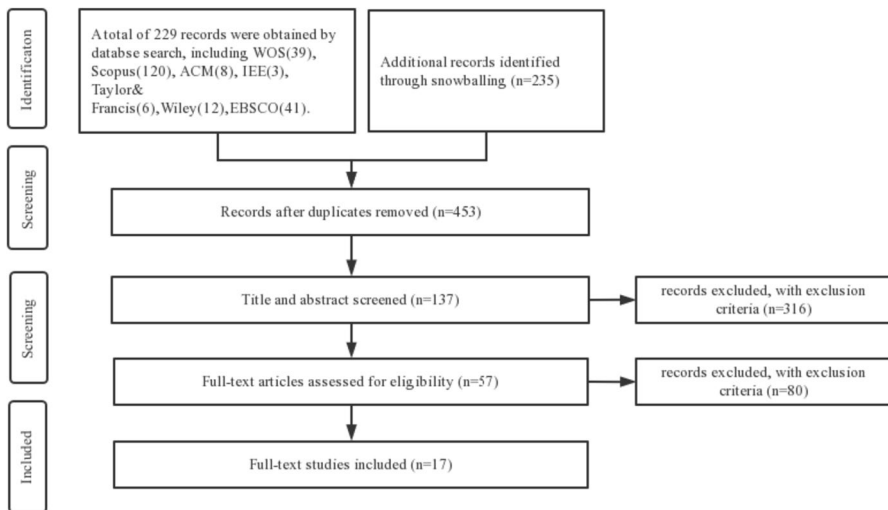


Fig. 1 PRISMA flow chart of selection process

application categories of AI-driven educational assessment in STEM education, we further provided detailed explanations of the categories and presented examples for corresponding categories (Hsieh & Shannon, 2005). The data analysis process was conducted by two researchers to ensure the reliability.

Results

Regarding the basic information of the 17 articles, most of them were published after the year of 2018 ($n=15$, 88.2%). The most frequent published country was China, followed by the USA and Spain. In addition, the 17 articles were published in 13 journals. Specifically, *Journal of Science Education and Technology* published the most articles ($n=3$, 17.6%), followed by *Computational Intelligence and Neuroscience* ($n=2$, 11.8%) and *International Journal of Artificial Intelligence in Education* ($n=2$, 11.8%).

RQ1: What Are the Functions of AI Applications in STEM Educational Assessment?

Among the 17 included studies, AI-driven educational assessment in STEM education was divided into three main functions: academic performance assessment ($n=12$, 70.6%), learning state assessment ($n=3$, 17.6%), and instructional quality assessment ($n=2$, 11.8%) (see Fig. 2).

Academic Performance Assessment

Among the 17 reviewed studies, 12 studies focused on assessing students' academic performance in STEM education. Two categorizations could be further revealed: formative assessment ($n=11$, 64.7%) and summative assessment ($n=1$, 5.9%). Specifically, in the first category of formative assessment, AI-driven educational assessment could help instructor evaluate students' formative learning performance and provided feedbacks. For example, Maestrales et al. (2021) designed an automated assessment tool for formative assessments during students' science learning. The results indicated that, compared to instructors' manual assessment, the machine learning-enabled assessment tool achieved more accuracy in scoring. Aiouni et al. (2016) proposed an automated grading system namely eAlgo that employed an

Fig. 2 The functions of AI-driven educational assessment in STEM education

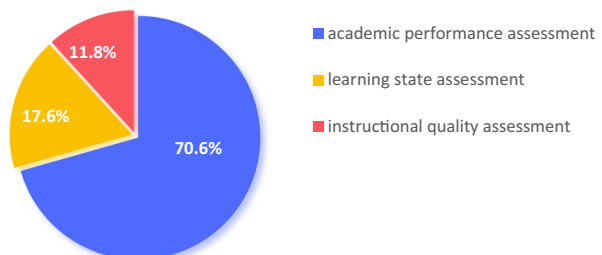


Fig. 3 The AI algorithms of AI-driven educational assessment in STEM education. Note: NLP, natural language processing; LO, logical operators; ML, machine learning; GM, graph matching; BN, Bayesian network; NF, neuro fuzzy; DL, deep learning; IPSO-BP, a BP neural network model based on improved particle swarm optimization. **a** The AI algorithms used in STEM educational assessment, **b** The distribution of AI algorithms use (by year), **c** The distribution of AI algorithms use (by functions)

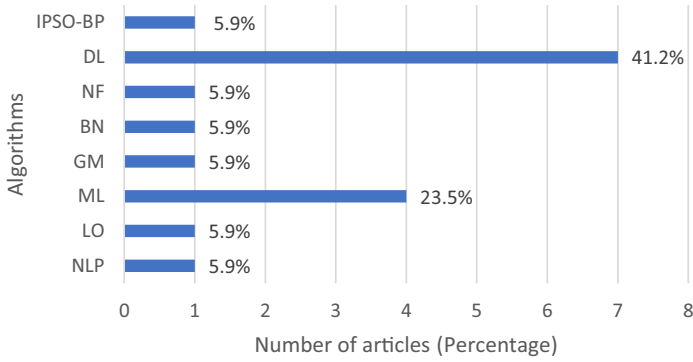
automated matching algorithm to assess students' flowcharts in the programming courses. Chen and Wang (2023) employed a deep learning evaluation model for formative assessment during the learning process of a programming course. In the second category of summative assessment, AI-driven educational assessment was used to evaluate students' final products and performance. For example, Zhai et al. (2021) utilized an effectiveness reasoning network to address the cognitive, instructional, and reasoning effectiveness issues in Next Generation Science Assessment (NGSA) based on machine learning, aiding instructors in scoring students' summative performance in science courses.

Learning State Assessment

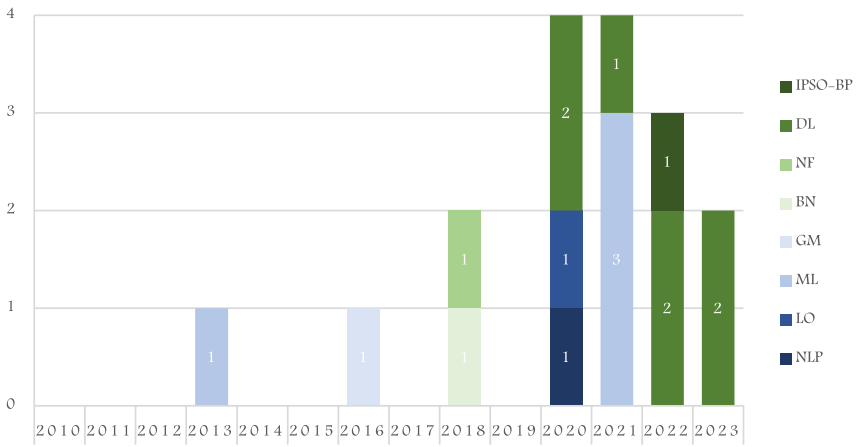
Among the 17 reviewed studies, 3 studies focused on assessing students' learning state in STEM education. With the support of AI-driven educational assessment, instructors can better understand students' learning states (e.g., emotional states, learning risks) during the STEM education. For example, Bertolini et al. (2021) combined a novel assessment method (i.e., concept maps) with machine learning techniques to assess the dropout risks of students in biology courses. Based on the assessment results, instructors could take actionable course-level interventions to help students. Zhang et al. (2022) combined AI-enabled facial recognition technology with actual instructional and management work in a smart classroom, to capture students' emotional states and changes during the STEM learning. Aljuaid and Said (2021) employed an effective deep learning-based method with a feedforward deep neural network classifier to assess students' learning effectiveness during the software engineering learning. Automatic assessment during the instructional process was conducted to help instructors track and evaluate students' learning status.

Instructional Quality Assessment

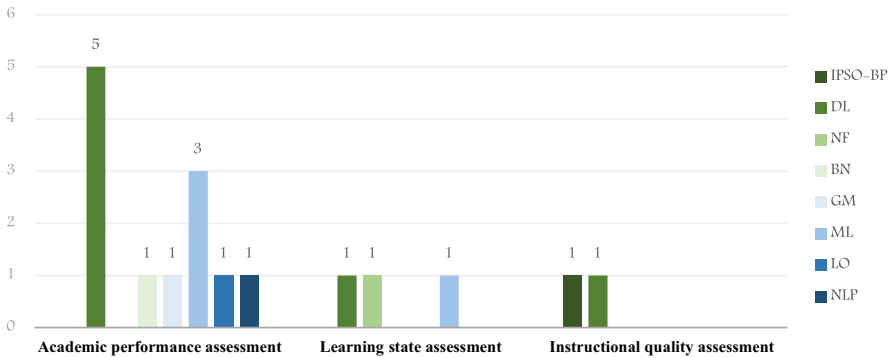
Among the 17 reviewed studies, 2 studies focused on valuating instructors' instructional quality in STEM education. Through assessing the instructional process, AI-driven educational assessment can help instructors modify the instructional strategies and improve instructional quality. For example, Chen et al. (2022) improved the particle swarm optimization algorithm to optimize the model structure and parameters of the IPSO-BP algorithm to provide a reliable educational assessment model for science course instruction. Furthermore, an assessment indicator system for instructional quality was established, including four main indicators and 16 sub-indicators, in order to quantify and evaluate the quality of instructional activities. Omer et al. (2020) employed DL algorithms to analyze students' cognitive maps and assessment data to assess their learning performance in a programming course. The



(a) The AI algorithms used in STEM educational assessment



(b) The distribution of AI algorithms use (by year)



(c) The distribution of AI algorithms use (by functions)

findings demonstrated that AI algorithms could assist instructors in gaining better understandings of students' cognitive levels during their STEM learning.

RQ2: What AI Algorithms Are Used to Achieve Those Functions in STEM Educational Assessment?

To further understand how to achieve the functions of STEM educational assessment, we clarified the AI algorithms used in STEM educational assessment (see Fig. 3a). It is worth noting that the AI algorithms used for comparison purposes were not analyzed in this research. Specifically, among the 17 reviewed studies, the most commonly used AI algorithms were deep learning (DL) ($n=7$, 41.2%), followed by machine learning (ML) ($n=4$, 23.5%) (see Fig. 3a). Other AI algorithms, such as natural language processing (NLP), logical operators (LO), graph matching (GM), Bayesian network (BN), neuro fuzzy (NF), and a BP neural network model based on improved particle swarm optimization (IPSO-BP), were also used in STEM educational assessment ($n=1$, 5.9%). Furthermore, we observed a decreasing trend in the frequency of traditional AI algorithms (e.g., ML, NLP) in the last several years. On the contrast, advanced AI algorithms (e.g., DL, NF) have been increasingly utilized in AI-driven educational assessment in STEM education (see Fig. 3b). In addition, the emergence of improved algorithms (e.g., IPSO-BP) has started to be applied in educational assessment of STEM education.

Specifically, this research found that multiple studies conducted AI-driven educational assessment in STEM education through traditional algorithms, including NLP, LP, ML, and GM (showed as blue colors in Fig. 3b). For example, Vittorini et al. (2021) developed a system using R commands that utilized NLP for automated assessment of students' data science assignments. The results indicated that the system not only reduced correction time and minimizes the possibility of errors but also provided automated feedback to support students in addressing exercise problems. Ariely et al. (2023) also employed ML algorithms to establish a reliable assessment model based on 500 examples to automatically evaluate students' writing assignments in science courses. Zhai et al. (2021) analyzed the potential validity issues of ML techniques in the assessment of science course learning, and the results revealed high consistency between the AI-driven evaluation and expert evaluation.

In the recent years, researchers have started to employ advanced AI algorithms such as DL, NF, BN, and IPSO-BP models in educational assessment of STEM education (showed as green colors in Fig. 3b). For example, Arfah Baharudin and Lajis (2021) used DL methods to assess students' performance in programming and identify obstacles encountered in the course, in order to motivate students to learn programming. He and Fu (2022) utilized DL algorithms and optimized instructional evaluation algorithms to enhance the quality of engineering course instruction. Deshmukh et al. (2018) conducted student educational assessment in network analysis courses using Mamdani fuzzy inference systems and neural fuzzy systems. The results indicated that using methods such as fuzzy and neural fuzzy systems based on classification criteria could achieve 100% classification accuracy in the assessment process. Chen et al. (2022) improved the efficiency and accuracy of

educational assessment in STEM learning by constructing an evaluation index system and optimizing the model structure and parameters using the IPSO-BP algorithm, which validated the reliability of the AI model.

Furthermore, the application of AI algorithms in three AI-driven education assessment functions was also clarified (see Fig. 3c). First, in the function of academic performance assessment, both advanced and traditional AI algorithms were utilized in educational assessment. A majority of studies employed DL in academic performance assessment, followed by ML and other algorithms (i.e., NLP, LP, GM, BN). For example, Lin and Chen (2020) developed a DL-based system to provide learning support and assess students' learning outcomes. Zhai et al. (2021) utilized traditional ML techniques for educational assessment, developing an assessment tool to evaluate students' progress in 3D learning. Second, in the function of learning state assessment, both traditional (i.e., ML) and advanced AI algorithms (i.e., DL, IPSO-BP) were used. For example, Zhang et al. (2022) assessed students' learning status through DL-based facial recognition to identify students' engagement during class. Third, in the function of instructional quality assessment, advanced AI algorithms, such as DL and IPSO-BP, were majorly used. For example, He and Fu (2022) assessed instructional quality by using advanced algorithms.

RQ3: What Are the Effects of AI Applications in STEM Educational Assessment?

To further understand the application effects of AI-driven educational assessment in STEM education, this research further clarified the educational and technological effects of AI applications in STEM educational assessment based on the empirical results and research conclusions of the 17 reviewed studies.

Educational Effects

First, among the 17 reviewed studies, 9 studies reported the positive impacts of AI-driven educational assessment on students' learning process. Specifically, the positive effects of AI-driven educational assessment on students included helping them overcome learning difficulties, providing practical knowledge strategies, identifying and intervening at-risk students, and improving the scientific validity and effectiveness of assessment outcomes. For example, Arfah Baharudin and Lajis (2021) assisted students with learning difficulties in a computer programming language course by using deep learning methods to identify their weaknesses and help them connect new and prior knowledge. Lin and Chen (2020) developed a deep learning recommendation system that incorporated augmented reality technology and learning theories to provide learning assistance to non-specialized students and students with diverse learning backgrounds. Bertolini et al. (2021) explored the role of novel assessment types and machine learning techniques to reduce the dropout rate of undergraduate science course, and the results indicated that the machine learning methods were helpful in addressing dropout issues. Ariely et al. (2023) proposed the use of natural language processing and deep learning techniques for automated

grading, which facilitated students to obtain assessment outcomes quickly and provided them with prompt feedback.

Second, among the 17 reviewed studies, 4 studies reported the positive impacts of AI-driven educational assessment on instructors' instructional process. Specifically, AI techniques were used in STEM educational assessment to help instructors identify instructional problems and improve assessment efficiency. For example, He and Fu (2022) proposed the application of deep learning methods in engineering courses to enhance the instructional quality. Through empirical validation, they demonstrated that the application of deep learning techniques in the evaluation of engineering course led to the improvement of instructional quality. Chen and Wang (2023) constructed a deep learning evaluation model and provided a new approach to improving the instructional process in programming courses, which could help instructors visualize the problem-solving patterns during students' programming learning. Overall, assessing the instructional process with the support of AI technologies can help instructors improve their instructional process and finally enhance the STEM education quality.

Third, the 17 reviewed studies mentioned various educational theories (e.g., three-dimensional learning theory, in-depth learning theory) to guide the application of.

Technological Effects

Among the 17 reviewed studies, 4 studies reported the technological benefits of using AI techniques in automated assessment in STEM education. Compared to traditional assessment in STEM education, AI-driven assessment can automatically collect and analyze data and improve the efficiency of assessment as well. In addition, the accuracy of formative and summative assessment in STEM education can be improved through AI-driven big data training and optimization. For instance, Vittorini et al. (2021) conducted a study on a new system that shortened the assessment time and provided automated feedback to support students in solving exercise problems during the STEM learning. The automated grading in this system showed an acceptable correlation with manual grading, and the system reduced grading time and helped instructors identify grading errors. Aiouni et al. (2016) proposed an automated process diagram algorithm grading system called eAlgo. In this AI-driven grading system, the identification of solutions was based on graph matching, where the method known as automatic matching of algorithm solutions could automatically assess students' algorithms based on pre-defined algorithms using similarity measurements as parameters.

Discussions and Implications

Addressing Research Questions

Recently, AI-driven educational assessment has been widely applied in STEM education, in order to empower the instructional and learning process. However, due

to the lack of relative review works, it is essential to delve into the applications of AI-driven educational assessment in STEM education to guide the future practice and research. Therefore, the current research conducted a systematic review to holistically examine the application of AI-driven educational assessment in STEM education, including the AI functions, algorithms, and effects. First, the results indicated the main AI functions of educational assessment in STEM education consisted of academic performance assessment, learning state assessment, and instructional quality assessment. Second, regarding the AI algorithms used in AI-driven educational assessment, the results revealed that the traditional AI algorithms (e.g., natural language processing, logical operators, machine learning, graph matching) as well as advanced AI algorithms (e.g., deep learning, neuro fuzzy, and IPSO-BP models) were applied. Compared to the traditional AI algorithms, the application of advanced AI algorithms was more prevalent in the last several years. Third, the effects of AI-driven educational assessment in STEM education were divided into educational effects and technological effects. Regarding the educational effects, AI-driven educational assessment could help students better understand their learning performance to further adjust their learning behaviors. Moreover, it could also help instructors reduce the assessment workload and enhance the instructional quality. Regarding the technological effects, compared to traditional educational assessment, AI-driven educational assessment had great potential to provide more automated, efficient, and accurate assessment in STEM education. Based on these findings, educational and technological implications were further proposed for future development of AI-driven educational assessment in STEM education.

Educational Implications

From the educational perspective, implications can be proposed from the perspectives of educational theories and educational processes. Regarding educational theories, the 17 reviewed studies mentioned various educational theories (e.g., three-dimensional learning theory, in-depth learning theory) and reached a consensus that achieving high-quality STEM education could not be accomplished by merely utilizing AI technologies (Castañeda & Selwyn, 2018; Du Boulay, 2000; Ouyang & Jiao, 2021; Selwyn, 2016). Specifically, one notable theory is the three-dimensional learning theory (Kaldaras et al., 2021), which emphasizes the needs for students to acquire interdisciplinary knowledge and skills to tackle challenges in the ever-changing domains of science and technology. Specifically, two studies in this review mentioned this theory when designing AI-driven educational assessment tools and refining the assessment process (i.e., Maestrales et al., 2021; Zhai et al., 2021). Under the guidance of the three-dimensional learning theory, educational assessment in STEM education highlighted to help students break down complex phenomena by integrating science and engineering practices, crosscutting concepts, and disciplinary core ideas to comprehend and address real-life problems. Another influential theory is the in-depth learning theory, which has received positive impacts from AI-driven educational assessment in STEM education. The in-depth learning theory shifts the research focus to student individuals, which highlights exploring learning from a

human-centered perspective. This theory advocates for students to study learning in authentic learning contexts, emphasizing a “problem-driven” paradigm in educational research. For instance, He and Fu (2022) determined the elements constituting project costs based on the in-depth learning theory and established a system of instructional evaluation indicators which focused on student individuals. However, more studies did not effectively integrate AI-driven educational assessment techniques with educational theories in STEM education. The lack of connections between educational theories and practical cases of AI-driven educational assessment in STEM education could be a possible reason. Recently, many researchers have proposed educational theories to guide the development and designs of AI techniques. For example, Ouyang and Jiao (2021) proposed three paradigms for AI in education (i.e., *AI-directed, learner-as-recipient, AI-supported, learner-as-collaborator*, and *AI-empowered, learner-as-leader*) from a theoretical perspective. These paradigms can serve as a reference framework for exploring different approaches to solve learning and instructional problems when using AI-driven assessment in STEM education. Compared to viewing learners as recipients and collaborators, future application of AI technologies in STEM educational assessment should be designed to empower learners to take full agency of learning and optimize AI technique to provide real-time insights about emergent learning.

Regarding educational processes, compared to using traditional assessment approaches, students might have a better understanding of their learning progress, and the instructors can focus more on the innovation of pedagogical designs. Specifically, with the support of AI-driven educational assessment, students can gain insights into their learning status in STEM education, enabling them to adjust and modify their learning behaviors. For instance, Zhang et al. (2022) applied deep learning-based facial recognition technology in smart classrooms to promote students’ learning process. In addition, the application of AI-driven educational assessment in STEM education might also alleviate the workload of instructors. In traditional STEM education, educational assessment relies on expert evaluation or self-assessment, which might reduce the efficiency of instruction and learning. However, with the emergence of AI techniques, the educational assessment process becomes more automatic and convenient. For example, Chen et al. (2022) improved the efficiency and accuracy of assessment by constructing an evaluation indicator system and optimizing the model structure and parameters using the IPSO-BP algorithm. Overall, the application of AI-driven educational assessment in STEM education has the potential to reshape the educational system through innovating educational theories and promoting educational processes.

Technological Implications

From the technological perspective, this research demonstrated the technical development of the application of AI-driven educational assessment in STEM education. Specifically, we found that previous researchers have utilized AI technologies and algorithms such as natural language processing, machine learning, genetic algorithms, as well as advanced algorithms like deep learning, neural fuzzy systems,

and IPSO-BP models to achieve automatic and efficient educational assessment in STEM education. By analyzing large amounts of students' learning performance data, AI algorithms can identify areas where students may be struggling and provide personalized recommendations for students (Zampirolli et al., 2021). Moreover, when training and evaluating large-scale data, AI technologies are able to produce more effective and accurate assessment outcomes than human beings (Ouyang et al., 2022). Moreover, compared to the traditional AI algorithms, we found that the application of advanced AI algorithms was more prevalent in STEM educational assessment in the past several years. Regarding the application of advanced AI algorithms, precise and effective assessment is the future direction for STEM educational assessment. Echoing this trend, future development of educational assessment is supposed to integrate with advanced AI algorithms and techniques (e.g., deep learning, neuro fuzzy, and IPSO-BP models). Zhang and Wang (2023) designed an intelligent knowledge discovery system for educational quality assessment using a genetic algorithm-based backpropagation neural network, and the results demonstrated system performance closed to human expert evaluations. Therefore, the application of AI technologies, as an increasing trend of educational assessment, provides more choices and flexibility for future technological development of STEM educational assessment.

In addition, AI-driven educational assessment has the potential to enhance the development of students' technical skills during STEM learning. Since AI-driven educational assessment involves the utilization and operation of various technological tools and platforms (e.g., machine learning models, smart classroom, intelligence tutoring system), students have more opportunities to develop their technology-related skills (e.g., computational thinking). Moreover, through interaction with the AI-enabled technologies, students can also improve their digital literacy, including data processing, algorithm comprehension, and application abilities. Hence, learning in AI-empowered environment will enhance the digital literacy of the next generation. Furthermore, the development of technical skills as well as digital literacy can help educators address issues related to biases in AI algorithms and the lack of transparency behind AI decision-making (Hwang & Tu, 2021; Hwang et al., 2020).

Conclusion, Limitations, and Future Direction

With the development of emerging technologies, AI techniques have been widely applied in educational assessment of STEM education. This research conducted a systematic review to provide an overview of the functions, algorithms, and effects of AI-driven educational assessment in STEM education. Specifically, we found that AI-driven educational assessment was primarily used to assess academic performance, learning status, and instructional quality. In addition, traditional algorithms (e.g., natural language processing, machine learning) as well as advanced algorithms (e.g., deep learning, neural fuzzy systems, IPSO-BP models) have been applied in STEM educational assessment. Moreover, the utilization of advanced AI-driven educational assessment has great potential to transform the educational and technological aspects of STEM education. Based on our findings, we proposed educational

and technological implications to guide future practice and development of AI-driven educational assessment in STEM education.

There are two main limitations in this systematic review, which lead to the future direction. First, we used the keywords related to AI-driven educational assessment in STEM education to locate the targeted studies. However, some specific AI algorithms or models might be not exhaustively searched, which cannot guarantee the integrity of searching results. Therefore, future review works are supposed to expand the range of keywords to avoid this problem. Second, the studies included in this review were limited to peer-reviewed journal articles. However, some advanced research about AI-driven educational assessment in STEM education may appear as conference papers or other forms; future systematic review can include these types of studies to further examine the application of AI-driven educational assessment in STEM education. Overall, this research provided educational practitioners and researchers an overview of AI-driven educational assessment, which has the potential to transform STEM education by providing efficient, objective, and data-driven evaluation methods.

All authors have approved the manuscript for submission, and the content of the manuscript has not been published, or submitted for publication elsewhere.

Funding This work was supported by National Natural Science Foundation of China (62177041); Zhejiang Province educational science and planning research project (2022SCG256); and Zhejiang University graduate education research project (20220310).

Data Availability The data is available upon request from the corresponding author.

Declarations

Conflict of Interest The authors declare no competing interests.

References

*Reviewed articles

- *Aiouni, R., Bey, A., & Bensebaa, T. (2016). An automated assessment tool of flowchart programs in introductory programming course using graph matching. *Journal of e-Learning and Knowledge Society*, 12(2). <https://www.learntechlib.org/p/173461/>
- Aljuaid, L. Z., & Said, M. Y. (2021). Deep learning-based method for prediction of software engineering project teamwork assessment in higher education. *Journal of Theoretical and Applied Information Technology*, 99(9), 2012–2030. <http://www.jatit.org/volumes/Vol99No9/8Vol99No9.pdf>
- *Arfah Baharudin, S., & Lajis, A. (2021). Deep learning approach for cognitive competency assessment in computer programming subject. *International Journal of Electrical and Computer Engineering Systems*, 1, 51–57. <https://hrcak.srce.hr/266730>
- *Ariely, M., Nazaretsky, T., & Alexandron, G. (2023). Machine learning and hebrew NLP for automated assessment of open-ended questions in biology. *International Journal of Artificial Intelligence in Education*, 33(1), 1–34. <https://doi.org/10.1007/s40593-021-00283-x>

- Bertolini, R., Finch, S. J., & Nehm, R. H. (2023). An application of Bayesian inference to examine student retention and attrition in the STEM classroom. *Frontiers in Education*, 8, 1073829. <https://doi.org/10.3389/educ.2023.1073829>
- *Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Testing the impact of novel assessment sources and machine learning methods on predictive outcome modeling in undergraduate biology. *Journal of Science Education and Technology*, 30, 193–209. <https://doi.org/10.1007/s10956-020-09888-8>
- Castañeda, L., & Selwyn, N. (2018). More than tools? Making sense of the ongoing digitizations of higher education. *International Journal of Educational Technology in Higher Education*, 15(1), 1–10. <https://doi.org/10.1186/s41239-018-0109-y>
- Chandio, M. T., Pandhiani, S. M., & Iqbal, R. (2016). Bloom's taxonomy: Improving assessment and teaching-learning process. *Journal of Education and Educational Development*, 3(2), 203–221. <https://doi.org/10.22555/joeeed.v3i2.1034>
- Chen, X. L., Zheng, J. Z., Du, Y. J., & Tang, M. W. (2020). Intelligent course plan recommendation for higher education: A framework of decision tree. *Discrete Dynamics in Nature and Society*, 2020, 7140797. <https://doi.org/10.1155/2020/7140797>
- *Chen, X., & Wang, X. (2023). Computational thinking training and deep learning evaluation model construction based on scratch modular programming course. *Computational Intelligence and Neuroscience*, 2023(3), 1–12. <https://doi.org/10.1155/2023/3760957>
- *Chen, L., Wang, L., & Zhang, C. (2022). Teaching quality evaluation of animal science specialty based on IPSO-BP neural network model. *Computational Intelligence and Neuroscience*, 2022, 3138885. <https://doi.org/10.1155/2022/3138885>
- Conley, D. (2015). A new era for educational assessment. *Education Policy Analysis Archives*, 23, 1–8. <https://files.eric.ed.gov/fulltext/ED559683.pdf>
- Debus, J. C., & Lawley, M. (2016). Benefits and drawbacks of computer-based assessment and feedback systems: Student and educator perspectives. *British Journal of Educational Technology*, 47(2), 294–301. <https://doi.org/10.1111/bjjet.12232>
- *Deshmukh, V., Mangalwede, S., & Rao, D. H. (2018). Student performance evaluation using data mining techniques for engineering education. *Advances in Science Technology and Engineering Systems Journal*, 3(6), 259–264. <https://doi.org/10.25046/aj030634>
- du Boulay, C. (2000). From CME to CPD: Getting better at getting better? Individual learning portfolios may bridge gap between learning and accountability. *BMJ*, 320(7232), 393–394. <https://doi.org/10.1136/bmj.320.7232.393>
- Ekolu, S. O. (2021). Model for predicting summative performance from formative assessment results of engineering students. *International Journal of Engineering Education*, 37(2), 528–536. https://www.ijee.ie/latestissues/Vol37-2/20_ijee4047.pdf
- Erickson, J. A., Botelho, A. F., McAteer, S., Varatharaj, A., & Heffernan, N. T. (2020). The automated grading of student open responses in mathematics. In C. Rensing, & H. Drachler (Eds.), *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 615–624). Association for Computing Machinery. <https://doi.org/10.1145/3375462.3375523>
- Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo? *Journal of Computer Assisted Learning*, 37(5), 1207–1216. <https://doi.org/10.1111/jcal.12577>
- *Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences*, 22(4), 521–563. <https://doi.org/10.1080/10508406.2013.837391>
- Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, 24(2), 105112. <https://doi.org/10.1016/j.nedt.2003.10.001>
- *He, X., & Fu, S. (2022). Data analysis and processing application of deep learning in engineering cost teaching evaluation. *Journal of Mathematics*, 2022(2), 1-12. <https://doi.org/10.1155/2022/8944570>
- Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, 48(8), 952–984. <https://doi.org/10.1002/tea.20439>
- Hendriks, W. J., Bakker, N., Pluk, H., de Brouwer, A., Wieringa, B., Cambi, A., Zegers, M., Wansink, D. G., Leunissen, R., & Klaren, P. H. (2019). Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores. *BMC Medical Education*, 19(1), 1–11. <https://doi.org/10.1186/s12909-019-1610-2>

- Holstein, K., McLaren, B. M., & Alevan, V. (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. *Artificial Intelligence in Education: 19th International Conference* (pp. 154–168). Springer International Publishing. https://doi.org/10.1007/978-3-319-93843-1_12
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Hwang, G. J., & Tu, Y. F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, 9(6), 1–19. <https://doi.org/10.3390/math9060584>
- Hwang, R. H., Peng, M. C., Huang, C. W., Lin, P. C., & Nguyen, V. L. (2020). An unsupervised deep learning model for early network traffic anomaly detection. *IEEE Access*, 8, 30387–30399. <https://doi.org/10.1177/30387-30399>
- Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial intelligence-enabled prediction model of student academic performance in online engineering education. *Artificial Intelligence Review*, 35, 6321–6344. <https://doi.org/10.1007/s10462-022-10155-y>
- Kaldaras, L., Akaeze, H., & Krajcik, J. (2021). Developing and validating Next Generation Science Standards-aligned learning progression to track three-dimensional learning of electrical interactions in high school physical science. *Journal of Research in Science Teaching*, 58(4), 589–618. <https://doi.org/10.1002/tea.21672>
- Kennedy, T. J., & Odell, M. R. (2014). Engaging students in STEM education. *Science Education International*, 25(3), 246–258. <https://files.eric.ed.gov/fulltext/EJ1044508.pdf>.
- *Lin, P. H., & Chen, S. Y. (2020). Design and evaluation of a deep learning recommendation based augmented reality system for teaching programming and computational thinking. *IEEE Access*, 8, 45689–45699. <https://doi.org/10.1109/ACCESS.2020.2977679>
- Lin, J. W., & Lai, Y. C. (2014). Using collaborative annotating and data mining on formative assessments to enhance learning efficiency. *Computer Applications in Engineering Education*, 22(2), 364–374. <https://doi.org/10.1002/cae.20561>
- *Maestrales, S., Zhai, X., Touitou, I., Baker, Q., Schneider, B., & Krajcik, J. (2021). Using machine learning to score multi-dimensional assessments of chemistry and physics. *Journal of Science Education and Technology*, 30, 239–254. <https://doi.org/10.1007/s10956-020-09895-9>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Neuendorf, K. A., & Kumar, A. (2015). Content analysis. *The International Encyclopedia of Political Communication*, 9781118541555, 1–10. <https://doi.org/10.1002/9781118541555.wbiepc065>
- *Omer, U., Farooq, M. S., & Abid, A. (2020). Cognitive learning analytics using assessment data and concept map: A framework-based approach for sustainability of programming courses. *Sustainability*, 12(17), 6990. <https://doi.org/10.3390/su12176990>
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers & Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27(6), 7893–7925. <https://doi.org/10.1007/s10639-022-10925-9>
- Saito, T., & Watanobe, Y. (2020). Learning path recommendation system for programming education based on neural networks. *International Journal of Distance Education Technologies*, 18(1), 36–64. <https://doi.org/10.1007/10.4018/IJDET.2020010103>
- Selwyn, N. (2016). Digital downsides: Exploring university students' negative engagements with digital technology. *Teaching in Higher Education*, 21(8), 1006–1021. <https://doi.org/10.1080/13562517.2016.1213229>
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157–167. <https://doi.org/10.1016/j.chb.2014.05.038>
- Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones II, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., Lowe, A., Newman, S., Okolo, V., Olroyd, S., Peacock, B. R., Pickett, S. B., Slager, D. L., Caviedes-Solis, I. W., Stanchak, K. E., Sundaravandan, V., Valdebenito, C., Williams, C. R., Zinsli, K., & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering,

- and math. *Proceedings of the National Academy of Sciences*, 117(12), 6476–6483. <https://doi.org/10.1073/pnas.1916903117>
- Tseng, C. H., Lin, H. C. K., Huang, A. C. W., & Lin, J. R. (2023). Personalized programming education: Using machine learning to boost learning performance based on students' personality traits. *Cogent Education*, 10(2), 2245637. <https://doi.org/10.1080/2331186X.2023.2245637>
- Vennix, J., den Brok, P., & Taconis, R. (2018). Do outreach activities in secondary STEM education motivate students and improve their attitudes towards STEM? *International Journal of Science Education*, 40(11), 1263–1283. <https://doi.org/10.1080/09500693.2018.1473659>
- *Vittorini, P., Menini, S., & Tonelli, S. (2021). An AI-based system for formative and summative assessment in data science courses. *International Journal of Artificial Intelligence in Education*, 31(2), 159–185. <https://doi.org/10.1007/s40593-020-00230-2>
- Wang, L., Xue, J., Zheng, N., & Hua, G. (2011). Automatic salient object extraction with contextual cue. *2011 international conference on computer vision* (pp. 105–112). IEEE. <https://doi.org/10.1109/ICCV.2011.6126231>
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In M. Shepperd, T. Hall, & I. Myrteit (Eds.), *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (pp. 1–10). Association for Computing Machinery. <https://doi.org/10.1145/2601248.2601268>
- Xing, W., Li, C., Chen, G., Huang, X., Chao, J., Massicotte, J., & Xie, C. (2021). Automatic assessment of students' engineering design performance using a Bayesian network model. *Journal of Educational Computing Research*, 59(2), 230–256. <https://doi.org/10.1177/0735633120960422>
- Yang, T.-C. (2023). Application of artificial intelligence techniques in analysis and assessment of digital competence in university courses. *Educational Technology & Society*, 26(1), 232–243. [https://doi.org/10.30191/ETS.202301_26\(1\).0017](https://doi.org/10.30191/ETS.202301_26(1).0017)
- Zampirolli, F. A., Borovina Josko, J. M., Venero, M. L., Kobayashi, G., Fraga, F. J., Goya, D., & Savegnago, H. R. (2021). An experience of automated assessment in a large-scale introduction programming course. *Computer Applications in Engineering Education*, 29(5), 1284–1299. <https://doi.org/10.1002/cae.22385>
- Zehner, F., & Hahnel, C. (2023). Artificial intelligence on the advance to enhance educational assessment: Scientific clickbait or genuine gamechanger? *Journal of Computer Assisted Learning*, 39(3), 695–702. <https://doi.org/10.1111/jcal.12810>
- *Zhai, X., Krajcik, J., & Pellegrino, J. W. (2021). On the validity of machine learning-based next generation science assessments: A validity inferential network. *Journal of Science Education and Technology*, 30, 298–312. <https://doi.org/10.1007/s10956-020-09879-9>
- *Zhang, X., Zhang, X., & Dolah, J. B. (2022). Intelligent classroom teaching assessment system based on deep learning model face recognition technology. *Scientific Programming*, 2022, 1851409. <https://doi.org/10.1155/2022/1851409>
- Zhang, H., & Wang, J. (2023). A smart knowledge discover system for teaching quality evaluation via genetic algorithm based BP neural network. *IEEE Access*, 11, 53615–53623. <https://doi.org/10.1109/ACCESS.2023.3280633>
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.