International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

INTERNATIONAL JOURNAL OF MODERN EDUCATION (IJMOE)
www.ijmoe.com

GLOBAL ACADEMIC EXCELLENCE

CODE, CLICK, LEARN: A SYSTEMATIC REVIEW OF ONLINE ASSESSMENT TOOLS IN 21ST CENTURY PROGRAMMING EDUCATION

Magendran Munisamy[1*], Siti Zuraidah Md Osman[2], Mageswaran Sanmugam[3]

[1] School of Education, Universiti Sains Malaysia, Malaysia
Email: tsmagendran@student.usm.my
[2] School of Education, Universiti Sains Malaysia, Malaysia
Email: sitizuraidah@usm.my
[3] Centre for Instructional Technology and Multimedia, Universiti Sains Malaysia, Malaysia
Email: mageswaran@usm.my
[*] Corresponding Author

Article Info:

Article history:
Received date: 10.01.2024
Revised date: 28.01.2024
Accepted date: 20.02.2024
Published date: 12.03.2024

To cite this document:

Munisamy, M., Osman, S. Z. M., & Sanmugam, M. (2024). Code, Click, Learn: A Systematic Review Of Online Assessment Tools In 21st Century Programming Education. *International Journal of Modern Education, 6* (20), 358-377.

DOI: 10.35631/IJMOE.620027

Abstract:

This study investigates the contemporary landscape of 21st-century programming education, recognizing the imperative for adaptive pedagogical approaches. Delving into the evolution and impact of automated assessment tools in programming education, the research identifies the problem of enhancing student proficiency through effective assessment methodologies. The dual purpose of the study is to elucidate diverse teaching methods and strategies, highlighting innovative approaches fostering comprehensive comprehension and skill acquisition and to unravel the transformative role of Technology-Enhanced Learning and Assessment Tools in Programming Education, exploring their implications on pedagogical practices and learning outcomes. Employing advanced search techniques, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodological framework uses advanced search techniques to meticulously analyse 27 scholarly articles obtained from the Web of Science, ERIC, and SCOPUS databases using keywords such as "programming," "online assessment," "tool," and "feedback," resulting in a synthesis that encapsulates the current state of programming education. The principal results underscore the efficacy of automated assessments in gauging student proficiency and illuminate diverse teaching methods fostering enhanced comprehension. It also highlights three pivotal themes: (1) Automated Assessment in Programming Education, (2) Teaching Methods and Strategies in Programming Education, and (3) Technology-Enhanced Learning and Assessment Tools in Programming Education. The major conclusions drawn from this comprehensive synthesis serve as a guide for teachers, policymakers, and researchers navigating the

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

dynamic intersection of technology and pedagogy in programming education, laying a robust foundation for future advancements in the field.

Keywords:

Programming; Online Assessment; Online Assessment Tool; Feedback

Introduction

The advent of online assessment tools has marked a revolutionary shift in programming education, altering traditional pedagogical dynamics and evaluation methodologies. The rapidly evolving digital landscape necessitates an educational paradigm that not only embraces technological advancements but also addresses the unique challenges and opportunities they present. Traditional evaluation methods in programming education are increasingly deemed insufficient due to their inability to cater to the dynamic needs of learners and educators in the 21st century (Barra et al., 2020). This inadequacy underscores a significant problem: the need for adaptive, efficient, and student-centered assessment methodologies that align with contemporary educational demands and learning styles (Choudhary et al., 2021). The literature substantiates the pressing need for innovative assessment tools. Studies by Gidvarowart et al. (2023), Karnalim et al. (2023) and Moosa & Bahaaudeen (2023) highlight the transformative potential of online assessment tools in enhancing learning outcomes, engagement, and instructional practices. Furthermore, the integration of Artificial Intelligence (AI) and Machine Learning (ML) within these tools offers unprecedented opportunities for personalized learning experiences, as detailed by Amer et al. (2021) and Surahman & Wang (2022).However, despite these advancements, the literature also points to a gap in comprehensive understanding and application of such technologies in programming education, necessitating further exploration.

This study sets its perimeter with clear objectives: to investigate the evolution and impact of automated assessment tools in programming education and to explore the pedagogical implications of diverse teaching methodologies facilitated by these technologies. Unique to this research is the systematic review of 27 scholarly articles using advanced search techniques, providing a synthesis that not only highlights the efficacy of automated assessments but also sheds light on innovative teaching methods that foster enhanced comprehension and skill acquisition.

By navigating the symbiotic relationship between technology and education, this article aims to comprehend the transformative power of "Code, Click, Learn" in sculpting the future of programming education. This exploration is not only timely but essential, as it addresses the critical gap identified in the literature by offering insights into the use of online assessment tools as both evaluative instruments and facilitators of an enriched learning experience.

Literature Review

The incorporation of online assessment tools into 21st-century programming education has significantly reshaped the educational landscape. Previous research has laid the groundwork for comprehending the evolution of these tools and their substantial influence on student learning outcomes, engagement, and instructional approaches. Earlier scholars, such as Louka (2022) and Parissi et al. (2023), made initial contributions by exploring the early stages of online assessment tools, providing a foundation for subsequent investigations. Expanding upon this foundation, a pivotal study conducted by Fernandez-Gauna et al. (2023) and Yan et al.

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

(2019) delved into the efficacy of online coding assessments. Their work underscored the importance of real-time feedback in improving student learning, offering insights into the crucial role of immediate guidance in programming skill development. In continuation of this discussion, Elmunsyah et al. (2022), Hsueh et al. (2023), Speth et al. (2022) and Venter (2022) investigated the incorporation of gamification elements in online programming tests. By examining how gamified features motivate students and contribute to a deeper understanding of coding concepts, these studies expanded our comprehension beyond traditional assessment methods.

In a parallel context, Lim et al. (2023) addressed the crucial aspect of accessibility in programming education. Their examination of asynchronous assessments conducted via online platforms highlighted the significance of catering to diverse learning styles, aligning with the broader theme of personalized learning experiences. Shifting focus to teacher perspectives, Anghelo Josué et al. (2023) and T. Gupta et al. (2023) illuminated how online assessment tools empower instructors to adapt their teaching methods. Furthermore, these studies explored how real-time analytics enable teachers to promptly identify and support students facing challenges, contributing significantly to discussions on data-driven decision-making and pedagogical adaptability.

Within the domain of technological integration, Hemachandran et al. (2024), Savelka et al. (2023), Smolansky et al. (2023) and Wermelinger (2023) delved into the integration of artificial intelligence (AI) into online programming assessments. Their research explored how AI-powered tools provide customized advice and recommendations, showcasing the potential for AI not only in evaluation but also in enhancing learning outcomes for a diverse range of learners.

Material and Methods

Systematic reviews are required to distil and outline the existing literature and provide a thorough and rigorous analysis that supports decision-making and guides future research. However, systematic reviews can differ in terms of quality and transparency, which may have an impact on their validity and usefulness. Thus, several reporting guidelines have been created to guarantee that systematic reviews are conducted and reported openly and consistently (Kim et al., 2021). To enhance the calibre and openness of this systematic review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement is applied to this work. Note that authors utilize the PRISMA statement as a tool, which is an evidence-based guide that includes a flowchart and checklist.

Identification

Several critical stages in the systematic review process were used to identify a large volume of relevant literature for this investigation. Initially, keywords were chosen, and then associated terms were investigated using dictionaries, thesauruses, encyclopaedias, and previous research (Azmi et al., 2024; Mat Sa'ud et al., 2023). Following the development of search queries for the WOS, ERIC, and SCOPUS databases, a list of relevant keywords was created, as shown in Table 1. In the first phase of the systematic review, 864 publications relevant to the current research project were successfully obtained via the two databases.

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Table 1: The Search String

SCOPUS	TITLE-ABS-KEY ((programming OR coding OR "computer science*") AND ("online assessment*" OR "on-line assessment*" OR "online formative assessment*" OR "on-line formative assessment*" OR "electronic formative assessment*" OR "e-assessment*" OR "eassessment*" OR "electronic assessment*" OR "flipped assessment*" OR "hybrid assessment*" OR "blended assessment*" OR "blended e-assessment" OR "blended electronic assessment*" OR "blended online assessment*" OR "blended eassessment" OR "authentic e-assessment*" OR "authentic electronic assessment*" OR "authentic online assessment*" OR "adaptive e-assessment*" OR "adaptive electronic assessment*" OR "adaptive online assessment*") AND (tool* OR system OR software OR application OR mechanism OR method)) AND (LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2023)) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))
ERIC	(programming OR coding OR "computer science*") AND ("online assessment*" OR "on-line assessment*" OR "online formative assessment*" OR "on-line formative assessment*" OR "electronic formative assessment*" OR "e-assessment*" OR "eassessment*" OR "electronic assessment*" OR "flipped assessment*" OR "hybrid assessment*" OR "blended assessment*" OR "blended e-assessment" OR "blended electronic assessment*" OR "blended online assessment*" OR "blended eassessment" OR "authentic e-assessment*" OR "authentic electronic assessment*" OR "authentic online assessment*" OR "adaptive e-assessment*" OR "adaptive electronic assessment*" OR "adaptive online assessment*") AND (tool* OR system OR software OR application OR mechanism OR method) (publicationtype: "Journal Articles" OR publicationtype: "Collected Works - Proceedings") language: English pubyearmin: 2020
Web Of Science	(programming OR coding OR "computer science*") AND ("online assessment*" OR "on-line assessment*" OR "online formative assessment*" OR "on-line formative assessment*" OR "electronic formative assessment*" OR "e-assessment*" OR "eassessment*" OR "electronic assessment*" OR "flipped assessment*" OR "hybrid assessment*" OR "blended assessment*" OR "blended e-assessment" OR "blended electronic assessment*" OR "blended online assessment*" OR "blended eassessment" OR "authentic e-assessment*" OR "authentic electronic assessment*" OR "authentic online assessment*" OR "adaptive e-assessment*" OR "adaptive electronic assessment*" OR "adaptive online assessment*") AND (tool* OR system OR software OR application OR mechanism OR method) (Topic) and 2020 or 2021 or 2022 or

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

2023 (Publication Years) and Proceeding Paper or Article (Document Types) and English (Languages)

Screening

In the initial phase, researchers established inclusion and exclusion criteria to screen 264 articles (refer to **Table 2**). During the second phase, 31 articles were excluded due to duplication. The primary criterion for selection was literature in the form of research articles and conference proceedings, as it provided valuable information. As a result, the study's scope excluded books, book chapters, meta-analyses, reviews, systematic reviews, and critiques. Furthermore, the review only considered papers written in English. Notably, the study focused on a four-year period (2020–2023). A specific criterion was applied to eliminate a total of 600 publications.

Table 2: The Criteria Used in the Search Selection

Criterion	Inclusion	Exclusion
Language	English	Non-English
Timeline	2020 – 2023	< 2020
Literature type	Journal (Article), Conference	Book, Review
Publication Stage	Final	In Press

Eligibility

During the third stage, known as eligibility assessment, a total of 233 articles were compiled. In this stage, we carefully examined the titles and main content of every article to confirm their adherence to the predefined inclusion criteria and their pertinence to the research goals of the present study. As a result, 206 papers, articles, and data were disqualified because they were out of the field, their titles had no significant relation to the study's goal, their abstracts had nothing to do with the goal, and empirical data did not support full-text access. As a result, there are now 27 articles remaining for the next evaluation.

Data Abstraction and Analysis

The present investigation used an integrative analytical methodology as its primary evaluation strategy, emphasising the meticulous scrutiny and integration of various research methods, particularly those that use quantitative approaches. This strategic inquiry was aimed at identifying pertinent themes and subthemes. The initial phase encompassed gathering data, signifying the onset of theme development. As illustrated in Figure 1, the researchers rigorously analysed a set of 27 publications to extract statements or materials relevant to the focal areas of the current study. An extensive assessment of prominent studies in online assessment within programming courses was carried out, entailing a detailed scrutiny of the used methodologies and an exploration of the resultant research outcomes. The primary investigator, in collaboration with co-authors, determined themes grounded in the empirical evidence pertaining to the study's context. During the data analysis phase, a detailed record was maintained, chronicling the analytical methods, viewpoints, challenges, and insights crucial for the interpretation of the data. In the concluding phase, a comparative analysis was conducted to identify any discrepancies in the development of themes. In cases of conceptual divergences, the team engaged in internal dialogues to address these issues. The themes identified underwent refinement to ensure uniformity. Significantly, two experts, one in educational technology and

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

the other in computer programming, oversaw the analysis's selection to ensure that the issues identified were relevant. This phase of expert review was essential in confirming domain validity and in assuring the clarity, relevance, and appropriateness of each identified subtheme.

International Journal of Modern Education
EISSN: 2637-0905
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

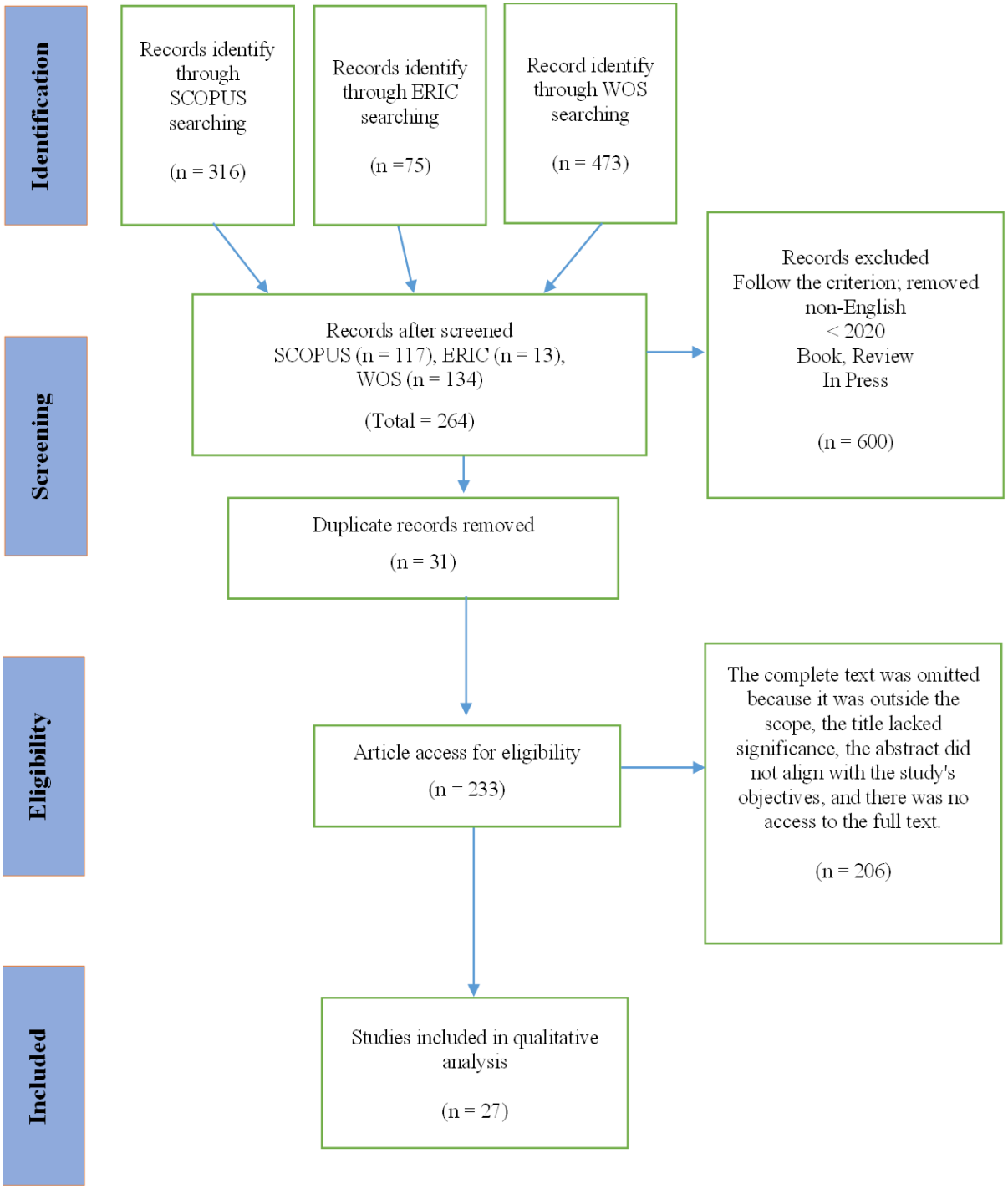

Figure 1: Diagram Outlining the Proposed Search Study Process (Page et al., 2021)

Result and Finding

Automated Assessment in Programming Education, Teaching Methods and Strategies in Programming Education, and Technology-Enhanced Learning and Assessment Tools in Programming Education are the three themes that group the 27 papers.

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Theme 1: Automated Assessment in Programming Education

Several studies address the challenges of assessing and supporting students in computer science education through innovative automated approaches. One approach involves a specialized computer program that analyses students' work, accurately grouping them based on performance and sentiment toward the material, eliminating the need for time-consuming surveys (Lokkila et al., 2022).

Another study focuses on testing programming skills, particularly in regions like Kosovo, emphasizing continuous learning and employing diverse teaching methods. This includes surveys, interviews, and a dedicated website for assessing computer skills (Jashari et al., 2023). To combat the shortage of computer experts, a collaborative effort creates a test using coding examples to automatically assess object-oriented programming skills and provide feedback for online courses with limited teacher guidance (Krugel et al., 2020; Satiman et al., 2024).

Acknowledging the widespread use of grading and feedback systems, this paper consolidates various systems to establish a comprehensive foundation for future research, enhancing the reliability, adaptability, security, and sustainability of educational assessment tools (Strickroth & Striewe, 2022). **Figure 2** demonstrates an interface that includes feedback. These research papers emphasise the ongoing evolution of automated assessment in the field of programming education. This includes the use of tools like Gradeer to evaluate coding and an investigation of the various factors that influence the effectiveness of online courses. It also introduces a rubric for standardized assessment, implements automated code-checking systems, and enhances student computer usage and programming understanding through tasks based on Bloom's taxonomy (Baranova & Simonova, 2021; Clegg et al., 2021; P. Gupta & Mehrotra, 2022; Insa et al., 2021; Sabjan et al., 2020).

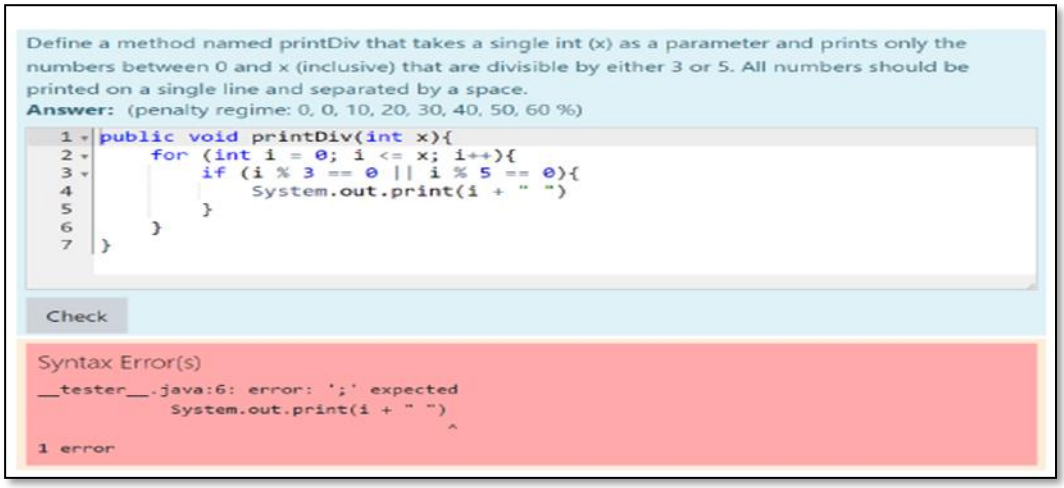

Figure 2: Example of Interface With Feedback (Caton et al., 2022)

Theme 2: Teaching Methods and Strategies in Programming Education

Several academic studies have been conducted to address the challenges posed by the COVID-19 pandemic, with a focus on pedagogical approaches and curricular strategies. One research paper explored the shifts in teaching approaches during the transition to online instruction, seeking feedback from students and teachers in Bahrain and Saudi Arabia (Moosa & Bahaaudeen, 2023). Another study focused on flipped learning, revealing its positive impact

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

on programming learning outcomes and emphasizing the utility of tests for targeted assistance (Cheng et al., 2021). In addition to addressing teaching evaluation, a novel expert system employed the internet and AI for real-time transmission and collection of multimedia monitoring information, showcasing its effectiveness in remote teaching assessment (Zhao, 2020). In response to the cheating challenge in programming assessments, the development of Dolos, a tool proficient in detecting code similarities, is discussed, contributing to fair assessments, even in online learning settings (Maertens et al., 2022). Additionally, Codeboard.io, an automated grading tool for programming assignments integrated with Measure of Software Similarity (MOSS) to address plagiarism, is presented, demonstrating successful testing in small and large classes (Appavoo & Meetoo-Appavoo, 2022). **Figure 3** illustrates one of technology-assisted teaching approach that uses Scratch for visual programming.

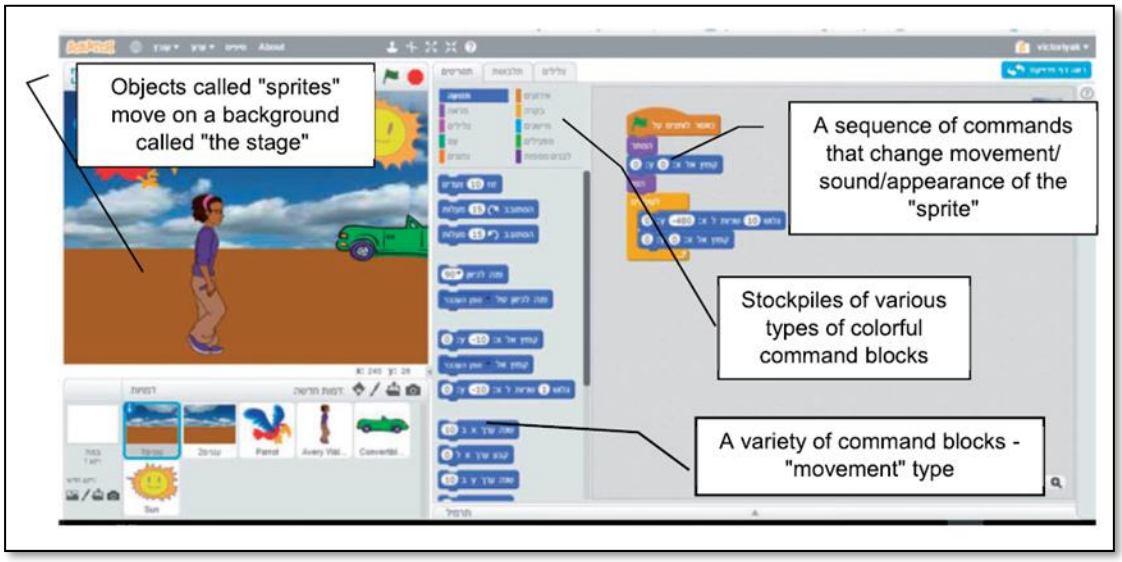

Figure 3: Visual Programming Using Scratch (Kesler et al., 2022)

Further studies explored diverse approaches, such as the use of an online tool called “rainy class,” revealing improved performance and engagement among less proficient students (Sun et al., 2020). Furthermore, the adaptation of teaching and testing methods during the global shift to online education highlights successful adjustments and the adaptability of teaching strategies (Muhammad & Srinivasan, 2021). Another study emphasized the significance of electronic assessments in enhancing student performance in Java programming classes over three years (Zietsman et al., 2020). Lastly, a study on gamification in programming education examined the impact of reward types on learning about plagiarism and cheating. This has revealed that grade-related rewards correlated with improved learning and engagement, albeit with a tendency for delayed assignment submissions (Karnalim et al., 2023). Collectively, these studies offer valuable insights into diverse teaching methods and tools, addressing the challenges and opportunities in the dynamic landscape of programming education.

International Journal of Modern Education
EISSN: 2637-0905
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Theme 3: Technology-Enhanced Learning and Assessment Tools in Programming Education

Technology-Enhanced Learning and Assessment tools have become increasingly important in programming education. They offer students interactive and captivating learning experiences, enabling them to practice coding in a simulated environment and receive immediate feedback on their efforts. As seen in **Figure 4**, the illustration of assessment tools for programming classification scheme, there is potential for further exploration of different combinations by researchers when developing new assessment tools. Additionally, it was observed that most tools do not have a specialty, with limited diversity noted among those that do specialize in particular approaches and assessment types.

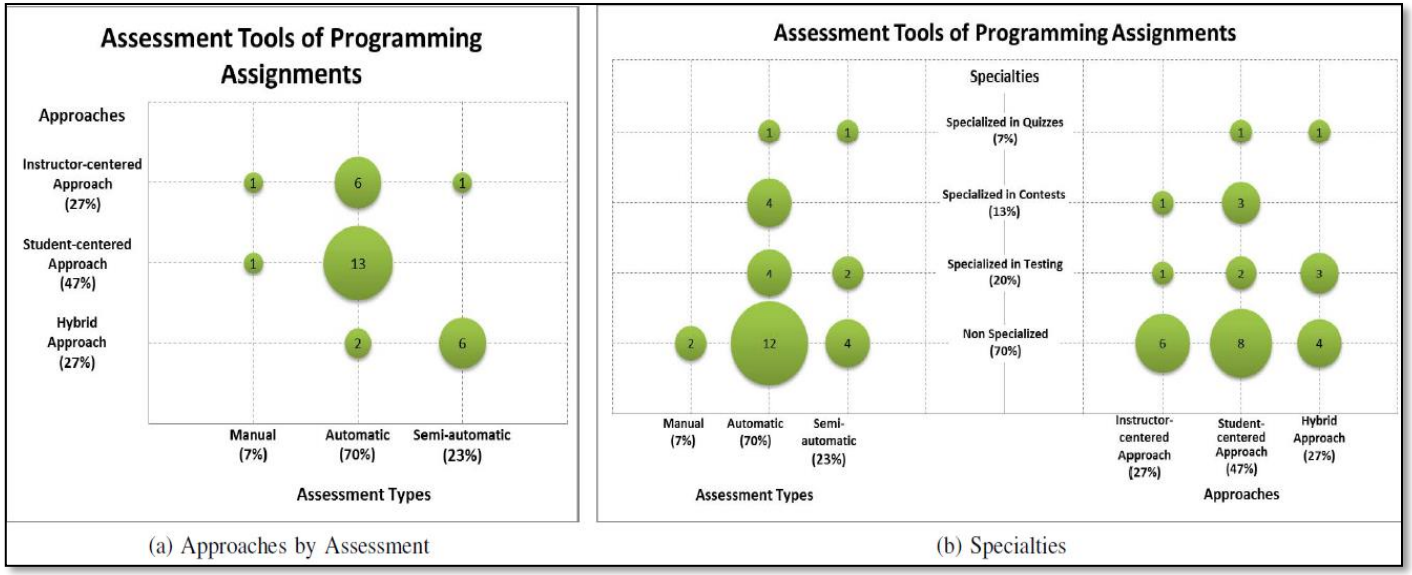

Figure 4: Example of Assessment Tools For Programming Classification Schemes (Souza et al., 2016)

Nine studies out of the 27 that were chosen to assess how technology can improve learning and assessment in programming education environments. **Table 3** summarize the research article findings based on Technology-Enhanced Learning and Assessment Tools in Programming Education theme.

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Table 3: The Research Article Finding Based on the Proposed Searching Criterion

Authors	Title	Source Title	Methodology	Findings and Advantages
Strickroth & Holzinger (2023)	Supporting the Semi-automatic Feedback Provisioning on Programming Assignments	Lecture Notes in Networks and Systems	This study investigated the role of teaching assistants in the semi-automated evaluation of programming assignments, with the goal of streamlining the process of providing effective feedback. The research involved improving an existing semi-automated electronic assessment system by incorporating customisable feedback snippets. Furthermore, the system included adaptively recommended feedback snippets derived from feedback on similar submissions. The goal was to assess the impact of these changes on grading efficacy and the nature of feedback provided by teaching assistants.	The study's findings revealed that using these feedback snippets resulted in more consistent and encouraging feedback. They also proved useful in detecting errors. Furthermore, it was observed that these snippets had no effect on the grading results.
Chrysafiadi et al. (2022)	A fuzzy-based mechanism for automatic personalized assessment in an e-learning system for computer programming	Intelligent Decision Technologies	The methodology of the study involved the use of specific criteria and fuzzy rules to enable the automatic personalized assessment of computer programming students in an e-learning environment. The real-world evaluation involved feedback from both students and experts to gauge the effectiveness of the presented mechanism.	The study's findings demonstrated the effectiveness of the fuzzy-based mechanism in creating personalized and balanced tests for computer programming students in an e-learning environment, as evidenced by the positive feedback from both students and experts.
Sherman et al. (2022)	Development of an electronic system for remote assessment of students' knowledge in cloud-based learning environment	CEUR Workshop Proceedings	The study employs systems analysis and methodological approaches to describe university departments through the lens of an invariant model of an organisation. The study also applied qualitative research methodology to assess and examine Knowledge Management (KM), Risk Management (RM), and Project Management (PM) during the undertaking of IT projects. Furthermore, the study looked at how the Delphi technique, a methodical approach to forecasting that uses panel experts' aggregated insights, was applied. This technique was	The findings emphasized the potential of the proposed system to enhance the efficiency of teacher time, motivate students to engage in honest learning, and contribute to the formation of an open information and cloud-based learning environment, particularly in the context of the rapid transformation of higher

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

			used to generate consensus among panel members, specifically in the field of healthcare research. The methodology aims to achieve homogeneity in the study and to address the specific needs of the educational environment, especially in the context of the rapid transformation of higher education brought about by the COVID-19 pandemic.	education driven by the COVID-19 pandemic.
Tkachuk et al. (2021)	Using Mobile ICT for Online Learning During COVID-19 Lockdown	Communications in Computer and Information Science	The investigation used a qualitative research methodology to develop and empirically validate approaches for incorporating mobile technologies into university students' education during the COVID-19 lockdown. This study intended to tailor mobile Information and Communication Technologies (ICT) for online pedagogy through an analysis of existing scholarly literature. The authors then created and examined a number of methods and systems, such as Audience Response Systems, Mobile Multimedia Authoring Tools, Mobile Learning Management Systems, Mobile Modelling and Programming Environments, and Mobile Database Management Systems. The study's methodological focus was on evaluating the functionality of these systems and empirically measuring the effectiveness of the developed technologies, with the goal of meeting the unique requirements of the educational landscape during the COVID-19 lockdown.	The study compared the capabilities of these five distinct systems and empirically supported the effectiveness of the developed technology. The findings support the efficacy of the proposed methodologies and demonstrate the viability of mobile Information and Communication Technologies (ICT) in facilitating online education during the COVID-19 lockdown period.
Pankiewicz (2020)	A warm-up for adaptive online learning environments - The Elo rating approach for assessing the cold start problem	ICCE 2020 - 28th International Conference on Computers in Education, Proceedings	The intent of the research is to present and assess the effectiveness of the Elo rating algorithm in determining the difficulty level of tasks, with a focus on the 'cold start' issue that arises during the early stages of deploying an adaptive system to users. The evaluation was carried out with actual data obtained from an interactive course delivered via the RunCode platform. This digital learning platform allows for multiple attempts and provides feedback following each submission. This analysis used a dataset of	Based on the study results, the Elo rating algorithm demonstrates a correlation coefficient of 0.702 with established reference values when the minimum sample size is n = 5. Furthermore, an improved correlation of 0.905 is observed with a sample size of n = 50. In comparison to the

International Journal of Modern Education
EISSN: 2637-0905
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

			50,055 submissions across 76 tasks contributed by 299 RunCode users.	Proportion Correct method, the Elo algorithm outperforms it for smaller sample sizes. As a result, it may be a more viable option as a simple technique for estimating task difficulty early in the development of an adaptive system for public use.
Marchisio et al. (2020)	Automatic Formative Assessment in Computer Science: Guidance to Model-Driven Design	Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020	The research paper advocated for the use of structured quality instruments for evaluating the Delphi method. This includes identifying key research issues, selecting appropriate panel members, ensuring panellist anonymity, effectively managing feedback, conducting iterative Delphi rounds, establishing consensus benchmarks, analysing consensus attainment, determining the criteria for completing the process, and evaluating the stability of the outcomes. The Delphi technique stresses the use of an expert panel as a means to establish uniformity within the study.	The study successfully implemented Online Adaptive Formative Education in Computer Science, specifically focusing on Model-Driven Design (MDD). The developed system incorporates an automatic formative assessment model with key features such as algorithmic questions, prompt feedback and unrestricted response formats. The transferability of these characteristics across academic disciplines allows the system to be expanded to include more subjects. The choice of MDD is significant due to its relevance to Computer Science education, particularly its connection with Computational Thinking, software design, and formal methods, which are areas requiring enhanced support.
Pereira et al. (2023)	Toward Human-AI Collaboration: A Recommender System to Support CS1 Teachers to	IEEE Transactions on Learning Technologies	The goal of this study is to help CS1 instructors choose problems for standardised or personalised assignments and exams by proposing an advanced AI-based recommender system. This system examines	Their recommendation system has an 88% accuracy rate, which is statistically significant ($p = 0.05$). Finally, these

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

	Select Problems for Assignments and Exams		student efforts within a POJ system's integrated development environment and automatically categorises CS1 problem topics based on their descriptions. Using data from 2714 students, the system helps teachers make better decisions. Its efficacy was tested against current standards in a blind experiment involving 35 CS1 teachers.	findings pave the way for the creation of innovative smart POJ learning environments in which educators can use AI technology to develop learning tasks such as homework and tests.
Smolansky et al. (2023)	Teacher and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education	L@S 2023 - Proceedings of the 10th ACM Conference on Learning @ Scale	A survey was conducted to explore teachers' and students' perspectives on innovative assessment practices, with a framework used to assess online review quality across six dimensions. The survey, which included 389 students and 36 teachers from two universities, revealed moderate use of generative AI, agreement on the types of assessments most impacted, and concerns about academic integrity. While teachers favoured assessments that incorporate AI and improve critical thinking, students had mixed feelings, partly due to concerns about reduced creativity.	The findings highlighted the importance of involving both teachers and students in assessment strategy reform, emphasising the prioritisation of learning processes over outcomes, the development of higher-order thinking skills, and the implementation of authentic applications.
Wang & Liang (2022)	CodingHere: Online Judge and Assessment System for Programming Course	5th IEEE Eurasian Conference on Educational Innovation 2022, ECEI 2022	The study adopts the Delphi technique, a systematic forecasting method using panel member consensus, widely recognized in medical fields. It proposes systematic quality tools for evaluating the Delphi method, covering aspects such as problem area identification, panel selection, panellist anonymity, controlled feedback, iterative rounds, consensus criteria, analysis, closing criteria, and result stability. Utilizing an expert panel, a key feature of the Delphi method, ensures uniformity in the study. The research study also stressed the importance of assessing the quality standards of Delphi studies. It provides critical evaluation metrics for researchers, editorial boards of medical journals, and reviewers. These metrics are intended to facilitate the evaluation of the quality of Delphi methodologies utilised for healthcare research.	The study's findings highlighted the effectiveness of CodingHere in supporting both teachers and students in the context of programming education, enabling efficient management of programming courses and providing students with valuable feedback to enhance their coding skills.

International Journal of Modern Education
IJMOE
EISSN: 2637-0905
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Discussion and Conclusion

The findings of this systematic review on automated assessment tools in programming education resonate with and extend upon existing research in several key areas. As highlighted in our study, the integration of technology in educational assessment has revolutionized programming education, echoing the observations of earlier studies (P. Gupta & Mehrotra, 2022; Lokkila et al., 2022).These advancements have facilitated a shift towards more personalized, efficient, and adaptive learning and assessment methods, aligning with the trends identified by (Gidvarowart et al., 2023; Karnalim et al., 2023). Additionally, our investigation highlights the significance of prompt feedback and the use of AI-powered tools to improve student learning outcomes, which aligns with the findings of Savelka et al. (2023) and Sherman et al. (2022). These tools not only improve the efficiency of assessments but also contribute significantly to the development of higher-order thinking skills, a critical aspect highlighted by (Smolansky et al., 2023) in their exploration of generative AI's impact on higher education assessments.

Additionally, the role of gamification elements in engaging students and fostering a conducive learning environment, as discussed in our review, finds support in the work of (Elmunsyah et al., 2022; Venter, 2022). The incorporation of these elements into programming education has been shown to enhance motivation and engagement, further emphasizing the need for innovative approaches to teaching and assessment in this field. Our findings also align with the broader implications of online assessment tools for pedagogical practices, as identified by (Moosa & Bahaaudeen, 2023). The transition to online and blended learning environments, accelerated by the COVID-19 pandemic, has underscored the versatility and adaptability of these tools in addressing diverse learning needs and scenarios.

In conclusion, our study contributes to the ongoing dialogue on the transformative potential of automated assessment tools in programming education. By drawing parallels with earlier research, we underscore the synergistic relationship between technological advancements and pedagogical innovation, highlighting the dynamic evolution of assessment practices in the digital age. The convergence of these tools with established educational theories and methodologies paves the way for a more inclusive, effective, and engaging programming education landscape.

Funding Statement

No financial support or grant funded this research.

Conflicts of Interest

The authors state that there are no conflicts of interest to disclose in the current study.

Acknowledgement

We sincerely thank all reviewers for their insightful and constructive feedback, as well as their collaboration and time commitment, which were critical to the successful completion of this review.

References

Amer, A., Alshehri, A., Saiari, H., Meshaikhis, A., & Alshamrany, A. (2021). Artificial Intelligence AI Assisted Thermography to Detect Corrosion under Insulation CUI. *SPE*

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Middle East Oil and Gas Show and Conference, MEOS, Proceedings, *2021-Novem*. https://doi.org/10.2118/204690-MS

Anghelo Josué, Bedoya-Flores, M. C., Mosquera-Quiñonez, E. F., Mesías-Simisterra, Á. E., & Bautista-Sánchez, J. V. (2023). Educational Platforms: Digital Tools for the teaching-learning process in Education. *Ibero-American Journal of Education & Society Research*, *3*(1). https://doi.org/10.56183/iberoeds.v3i1.626

Appavoo, P., & Meetoo-Appavoo, A. (2022). eExam Framework for Programming Classes. *Proceedings - 3rd International Conference on Next Generation Computing Applications, NextComp 2022*. https://doi.org/10.1109/NextComp55567.2022.9932218

Azmi, N. H., Ibrahim, N., Saari, E. M., Mutalib, A. A., & Akma, N. (2024). Student Profiling for Online Learning During Covid-19: A Systematic Review. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *34*(2), 50–61. https://doi.org/10.37934/araset.34.2.5061

Baranova, E., & Simonova, I. (2021). Taxonomy of learning objectives for the development of competencies ofcomputer sience teachers in a developing educational environment. *CEUR Workshop Proceedings*, *2920*, 8–19.

Barra, E., López-Pernas, S., Alonso, A., Sánchez-Rada, J. F., Gordillo, A., & Quemada, J. (2020). Automated Assessment in Programming Courses: A Case Study during the COVID-19 Era. *Sustainability (Switzerland)*, *12*(18), 1–24. https://doi.org/10.3390/SU12187451

Caton, S., Russell, S., & Becker, B. A. (2022). What Fails Once, Fails Again. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, 955–961. https://doi.org/10.1145/3478431.3499419

Cheng, S.-C., Cheng, Y.-P., Huang, Y.-M., & Yang, Y. (2021). Combining Flipped Learning and Formative Assessment to Enhance the Learning Performance of Students in Programming. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 13117 LNCS*. https://doi.org/10.1007/978-3-030-91540-7_51

Choudhary, F. R., Ahmed, S. Z., Sultan, S., & Khushnood, S. (2021). Comparative study of 21st Century Skills of Science Teachers and Students of Formal and Non-Formal Educational Institutes. *Review of Education, Administration & LAW*, *4*(1), 231–241. https://doi.org/10.47067/real.v4i1.131

Chrysafiadi, K., Virvou, M., & Tsihrintzis, G. A. (2022). A fuzzy-based mechanism for automatic personalized assessment in an e-learning system for computer programming. *Intelligent Decision Technologies*, *16*(4), 699–714. https://doi.org/10.3233/IDT-220227

Clegg, B., Villa-Uriol, M.-C., McMinn, P., & Fraser, G. (2021). Gradeer: An Open-Source Modular Hybrid Grader. *Proceedings - International Conference on Software Engineering*, 60–65. https://doi.org/10.1109/ICSE-SEET52601.2021.00015

Elmunsyah, H., Wibawa, A. P., Suswanto, H., Hidayat, W. N., Dwiyanto, F. A., Chandra, J. A., & Utomo, M. (2022). Online Programming Course Based on Gamification for First-Year Informatics Students. *JOURNAL OF ALGEBRAIC STATISTICS*, *13*(3).

Fernandez-Gauna, B., Rojo, N., & Graña, M. (2023). Automatic feedback and assessment of team-coding assignments in a DevOps context. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00386-6

Gidvarowart, S., Suchato, A., Wanvarie, D., Pratanwanich, N., & Tuaycharoen, N. (2023). Automated API Testing with Karate Framework: A Case Study of an Online

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Assessment Web Application. *Proceedings of JCSSE 2023 - 20th International Joint Conference on Computer Science and Software Engineering*, 309–314. https://doi.org/10.1109/JCSSE58229.2023.10202050

Gupta, P., & Mehrotra, D. (2022). Objective Assessment In JAVA Programming Language Using Rubrics. *Journal of Information Technology Education: Innovations in Practice*, *21*, 155–173. https://doi.org/10.28945/5040

Gupta, T., Shree, A., Chanda, P., & Banerjee, A. (2023). Online assessment techniques adopted by the university teachers amidst COVID-19 pandemic: A case study. *Social Sciences and Humanities Open*, *8*(1). https://doi.org/10.1016/j.ssaho.2023.100579

Hemachandran, V. C., Kumar, K. A., Sikandar, S. A., Sabharwal, S., & Kumar, S. A. (2024). A study on the impact of artificial intelligence on talent sourcing. *IAES International Journal of Artificial Intelligence*, *13*(1), 1 – 8. https://doi.org/10.11591/ijai.v13.i1.pp1-8

Hsueh, N. L., Xuan, Z. H., & Daramsenge, B. (2023). Design and Implementation of Gamified Learning System for Mutation Testing. *International Journal of Information and Education Technology*, *13*(7). https://doi.org/10.18178/ijiet.2023.13.7.1916

Insa, D., Pérez, S., Silva, J., & Tamarit, S. (2021). Semiautomatic generation and assessment of Java exercises in engineering education. *Computer Applications in Engineering Education*, *29*(5), 1034–1050. https://doi.org/10.1002/cae.22356

Jashari, X., Fetaji, B., & Guetl, C. (2023). Assessment of Digital Programing Skills based on the Competencies Model. *Proceedings - 2023 International Conference on Computing, Electronics and Communications Engineering, ICCECE 2023*, 157–162. https://doi.org/10.1109/iCCECE59400.2023.10238536

Karnalim, O., Simon, & Chivers, W. (2023). Non-game Incentives in Gamified Programming Education: More Marks or Prizes. *Lecture Notes in Networks and Systems*, *633 LNNS*. https://doi.org/10.1007/978-3-031-26876-2_86

Kesler, A., Shamir-Inbal, T., & Blau, I. (2022). Active Learning by Visual Programming: Pedagogical Perspectives of Instructivist and Constructivist Code Teachers and Their Implications on Actual Teaching Strategies and Students' Programming Artifacts. *Journal of Educational Computing Research*, *60*(1), 28–55. https://doi.org/10.1177/07356331211017793

Kim, M. M., Pound, L., Steffensen, I., & Curtin, G. M. (2021). Reporting and methodological quality of systematic literature reviews evaluating the associations between e-cigarette use and cigarette smoking behaviors: a systematic quality review. *Harm Reduction Journal*, *18*(1), 1–13. https://doi.org/10.1186/s12954-021-00570-9

Krugel, J., Hubwieser, P., Goedicke, M., Striewe, M., Talbot, M., Olbricht, C., Schypula, M., & Zettler, S. (2020). Automated measurement of competencies and generation of feedback in object-oriented programming courses. *IEEE Global Engineering Education Conference, EDUCON*, *2020-April*, 329–338. https://doi.org/10.1109/EDUCON45650.2020.9125323

Lim, R. S., Politz, J. G., & Minnes, M. (2023). Stream Your Exam to the Course Staff: Asynchronous Assessment via Student-Recorded Code Trace Videos. *SIGCSE 2023 - Proceedings of the 54th ACM Technical Symposium on Computer Science Education*, *1*. https://doi.org/10.1145/3545945.3569803

Lokkila, E., Christopoulos, A., & Laakso, M.-J. (2022). A Clustering Method to Detect Disengaged Students from Their Code Submission History. *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, *1*, 228–234. https://doi.org/10.1145/3502718.3524754

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Louka, K. (2022). Programming environments for the development of CT in preschool education: A systematic literature review. *Advances in Mobile Learning Educational Research*, *3*(1), 525–540. https://doi.org/10.25082/amler.2023.01.001

Maertens, R., Van Petegem, C., Strijbol, N., Baeyens, T., Jacobs, A. C., Dawyndt, P., & Mesuere, B. (2022). Dolos: Language-agnostic plagiarism detection in source code. *Journal of Computer Assisted Learning*, *38*(4), 1046–1061. https://doi.org/10.1111/jcal.12662

Marchisio, M., Margaria, T., & Sacchet, M. (2020). Automatic Formative Assessment in Computer Science: Guidance to Model-Driven Design. *Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020*, 201–206. https://doi.org/10.1109/COMPSAC48688.2020.00035

Mat Sa'ud, A., Md. Ghalib, Mohd. F., & Abu Bakar, R. (2023). A Structured Review Of Mobile Augmented Reality For Language Instruction And Learning. *International Journal of Modern Education*, *5*(19), 78–98. https://doi.org/10.35631/IJMOE.519006

Moosa, J., & Bahaaudeen, A. (2023). Programming courses Teaching methods Before, During, and After COVID-19 Pandemic. *2023 International Conference on IT Innovation and Knowledge Discovery, ITIKD 2023*. https://doi.org/10.1109/ITIKD56332.2023.10099736

Muhammad, N., & Srinivasan, S. (2021). Online education during a pandemic - adaptation and impact on student learning. *International Journal of Engineering Pedagogy*, *11*(3), 71–83. https://doi.org/10.3991/IJEP.V11I3.20449

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Pankiewicz, M. (2020). A warm-up for adaptive online learning environments - The Elo rating approach for assessing the cold start problem. *ICCE 2020 - 28th International Conference on Computers in Education, Proceedings*, *1*, 324–329.

Parissi, M., Komis, V., Dumouchel, G., Lavidas, K., & Papadakis, S. (2023). How Does Students' Knowledge About Information-Seeking Improve Their Behavior in Solving Information Problems? *Educational Process: International Journal*, *12*(1), 117–141. https://doi.org/10.22521/edupij.2023.121.7

Pereira, F. D., Rodrigues, L., Henklain, M. H. O., Freitas, H., Oliveira, D. F., Cristea, A. I., Carvalho, L., Isotani, S., Benedict, A., Dorodchi, M., & Oliveira, E. H. T. de. (2023). Toward Human–AI Collaboration: A Recommender System to Support CS1 Instructors to Select Problems for Assignments and Exams. *IEEE Transactions on Learning Technologies*, *16*(3), 457–472. https://doi.org/10.1109/TLT.2022.3224121

Sabjan, A., Abd Wahab, A., Ahmad, A., Ahmad, R., Hassan, S., & Wahid, J. (2020). MOOC Quality Design Criteria for Programming and Non-Programming Students. *Asian Journal of University Education*, *16*(4), 61–70. https://doi.org/10.24191/ajue.v16i4.11941

Satiman, L. H., Zulkifli, N., & Usman, A. (2024). Utilizing Online Quiz Assessment Tool to Provide Timely, Guided Feedback During COVID-19 Pandemic. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *35*(1), 88–96. https://doi.org/10.37934/araset.34.3.8896

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Savelka, J., Agarwal, A., An, M., Bogart, C., & Sakr, M. (2023). Thrilled by Your Progress! Large Language Models (GPT-4) No Longer Struggle to Pass Assessments in Higher Education Programming Courses. *ICER 2023 - Proceedings of the 2023 ACM Conference on International Computing Education Research V.1*, 78 – 92. https://doi.org/10.1145/3568813.3600142

Sherman, M. I., Samchynska, Y. B., & Kobets, V. M. (2022). Development of an electronic system for remote assessment of students' knowledge in cloud-based learning environment. *CEUR Workshop Proceedings*, *3085*, 290–305.

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. *L@S 2023 - Proceedings of the 10th ACM Conference on Learning @ Scale*, 378–382. https://doi.org/10.1145/3573051.3596191

Souza, D. M., Felizardo, K. R., & Barbosa, E. F. (2016). A Systematic Literature Review of Assessment Tools for Programming Assignments. *2016 IEEE 29th International Conference on Software Engineering Education and Training (CSEET)*, 147–156. https://doi.org/10.1109/CSEET.2016.48

Speth, S., Krieger, N., Reißner, G., & Becker, S. (2022). Teaching during the Covid-19 Pandemic - Online Programming Education. *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft Fur Informatik (GI)*, *P-321*. https://doi.org/10.18420/SEUH2022_09

Strickroth, S., & Holzinger, F. (2023). Supporting the Semi-automatic Feedback Provisioning on Programming Assignments. In *Lecture Notes in Networks and Systems: Vol. 580 LNNS*. https://doi.org/10.1007/978-3-031-20617-7_3

Strickroth, S., & Striewe, M. (2022). Building a Corpus of Task-Based Grading and Feedback Systems for Learning and Teaching Programming. *International Journal of Engineering Pedagogy*, *12*(5), 26–41. https://doi.org/10.3991/ijep.v12i5.31283

Sun, Q., Song, Y., & Tan, H. (2020). How do early programmers benefit from SPOC blended teaching: A data-driven analysis. *ACM International Conference Proceeding Series*, 65–70. https://doi.org/10.1145/3393527.3393539

Surahman, E., & Wang, T. H. (2022). Academic dishonesty and trustworthy assessment in online learning: A systematic literature review. *Journal of Computer Assisted Learning*, *38*(6). https://doi.org/10.1111/jcal.12708

Tkachuk, V., Yechkalo, Y., Semerikov, S., Kislova, M., & Hladyr, Y. (2021). Using Mobile ICT for Online Learning During COVID-19 Lockdown. In *Communications in Computer and Information Science* (Vol. 1308). https://doi.org/10.1007/978-3-030-77592-6_3

Venter, M. (2022). Online programming learning platform: The influence of gamification elements. *2022 IEEE IFEES World Engineering Education Forum - Global Engineering Deans Council, WEEF-GEDC 2022 - Conference Proceedings*. https://doi.org/10.1109/WEEF-GEDC54384.2022.9996263

Wang, J.-Y., & Liang, J.-C. (2022). CodingHere: Online Judge and Assessment System for Programming Course. *5th IEEE Eurasian Conference on Educational Innovation 2022, ECEI 2022*, 126–129. https://doi.org/10.1109/ECEI53102.2022.9829512

Wermelinger, M. (2023). Using GitHub Copilot to Solve Simple Programming Problems. *SIGCSE 2023 - Proceedings of the 54th ACM Technical Symposium on Computer Science Education*, *1*, 172 – 178. https://doi.org/10.1145/3545945.3569830

International Journal of Modern Education
EISSN: 2637-0905
IJMOE
Volume 6 Issue 20 (March 2024) PP. 358-377
DOI: 10.35631/IJMOE.620027

Yan, L., Hu, A., & Piech, C. (2019). Pensieve: Feedback on coding process for novices. *SIGCSE 2019 - Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. https://doi.org/10.1145/3287324.3287483

Zhao, X. (2020). Design of Teaching Expert Evaluation System Based on Artificial Intelligence. *Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Information Systems, ICAIIS 2020*, 675–679. https://doi.org/10.1109/ICAIIS49377.2020.9194904

Zietsman, E., Swart, K., & Daramola, O. (2020). Reflecting on e-assessment practices and students' performance in a Java programming course. In C. Busch, M. Steinicke, & T. Wendler (Eds.), *19th European Conference on e-Learning, ECEL 2020* (Vols. 2020-October, pp. 537–544). Academic Conferences and Publishing International Limited. https://doi.org/10.34190/EEL.20.012