Eurasian Journal of Educational Research 106 (2023) 321-337

Eurasian Journal of Educational Research
www.ejer.com.tr

The Dawn of ChatGPT: Transformation in Science Assessment

Suzan Alabidi[1], Khaleel AlArabi[2], Najeh Rajeh Alsalhi[3*], Maitha al Mansoori[4]

ARTICLE INFO

Article History:
Received: 13 May 2023
Received in revised form: 14 December 2023
Accepted: 08 January 2024
DOI: 0.14689/ejer.2023.106.019

Keywords
ChatGPT, Authentic Assessment, Traditional Assessment, Science Assessment.

ABSTRACT

The integration of Artificial Intelligence (AI) and ChatGPT in assessment has brought about a notable transformation in science education. The present study has undertaken an analysis of the limitations inherent in conventional assessments rooted in behaviourism while emphasising the significance of authentic assessments grounded in cognitivist principles. Furthermore, the utilisation of ChatGPT has demonstrated its capacity as a valuable instrument for the creation of interactive, customised, and captivating evaluation opportunities within the realm of science education. Educators have the potential to adopt AI technology in order to facilitate a transformative departure from passive recall-based assessments towards interactive and immersive learning experiences. The utilisation of ChatGPT's simulated interactions, individualised feedback, and adaptive assessment facilitates students' ability to actively participate in authentic scientific inquiry, employ critical thinking skills, and engage in problem-solving activities. In light of the evolving educational landscape, it is imperative to undertake a thorough examination and exploration of the potential impact of artificial intelligence (AI) on the future of science assessment.

[1] College of Education, Humanities and Social Sciences, Al Ain University. Email: suzan.alabidi@aau.ac.ae
[2] College of Education, Humanities and Social Sciences, Al Ain University. Email: Khaleel.alarabi@aau.ac.ae
[3] College of Arts, Humanities, and Social Sciences, University of Sharjah, Sharjah, U.A.E. Email: nalsalhi@sharjah.ac.ae
[4] College of Education, United Arab Emirates University, Al Ain, UAE. Email: 199902785@uaeu.ac.ae
* Corresponding Author Detail: Email: Najeh Rajeh Alsalhi; Email: nalsalhi@sharjah.ac.ae;
Mobile Number: 00971543374987

Introduction

The field of education is undergoing rapid transformation. The ongoing emergence of novel technologies and methodologies has led to a growing emphasis on the assessment of student learning. This matter is particularly urgent in the realm of science education. It is imperative to conduct a thorough analysis to ascertain whether our assessment methodologies effectively equip learners for the multifaceted challenges of the 21st century. It is possible that this may not be the case. However, the implementation of innovative technologies such as ChatGPT has the potential to revolutionise the field of science assessments, making them more authentic, engaging, and in line with current pedagogical approaches.

In the field of physics, examinations have traditionally emphasised the rote memorization and regurgitation of information, influenced by behaviourist principles that prioritise observable and quantifiable results (Quansah, 2018). Although the administration of conventional tests is relatively simple, they frequently fall short in accurately assessing comprehension, critical thinking, and the practical application of knowledge. The increasing prevalence of student-centred, cognitive learning approaches necessitates a critical re-evaluation of traditional science assessments (van Wyngaarden, Leech, & Coetzee, 2019). Educational professionals necessitate novel resources that facilitate dynamic assessments that are congruent with interactive pedagogical approaches, thereby equipping students with the necessary skills for thriving in the digital era.

One notable advancement in the field of artificial intelligence (AI) is ChatGPT, a chatbot that exhibits a high degree of human-like conversational abilities. ChatGPT, an AI system created by Anthropic, employs advanced machine learning techniques to produce individualised and contextually informed replies (Adiwardana et al., 2020). Preliminary results suggest that ChatGPT exhibits promising capabilities in evaluating science learners through interactive means, delivering personalised feedback, and facilitating immersive, student-centric experiences that align effectively with contemporary pedagogical approaches. Nevertheless, there is a lack of scholarly understanding regarding the extent to which ChatGPT is capable of evaluating scientific content. Therefore, this study aims to address this gap and make important contributions by investigating:

RQ1: How can ChatGPT overcome limitations of traditional assessments and support more authentic, cognitivist models in science?

RQ2: What specific ChatGPT assessment strategies can science educators employ?

RQ3: What are the implications of ChatGPT-based assessments for student engagement, learning, and outcomes?

RQ4: What challenges may arise in integrating ChatGPT, and how can they be mitigated?

This study undertakes a comprehensive examination of existing scholarly works that explore the convergence of science assessment, educational technology, and artificial intelligence. By synthesising the findings from these sources, the study establishes a set of principles that are grounded in empirical evidence, aiming to guide the optimal utilisation of ChatGPT. The proposal presents practical recommendations that are specifically designed to meet the needs of science teachers. This study emphasises the transformative potential of ChatGPT as an assessment tool that aligns with established theories. It

envisions a future where AI facilitates authentic and personalised science evaluations, thereby engaging learners and enabling them to reach their maximum potential.

The study holds significance for science education across multiple dimensions. The aforementioned statement highlights the potential of ChatGPT in facilitating enhanced assessment, which is a crucial aspect in the advancement of science education. The research advocates for the implementation of novel artificial intelligence (AI) methods that foster enhanced understanding, analytical reasoning, and active involvement. The focus lies on the integration of technology, such as ChatGPT, to facilitate the development of customised assessment experiences that enhance outcomes and preparedness for the digital era.

Moreover, this study establishes a connection between theory and practice by illustrating how assessments conducted using ChatGPT are consistent with essential learning principles and facilitate the effective evaluation of students' scientific knowledge and skills. The utilisation of advances in artificial intelligence (AI) in science assessment practices is encouraged, as it facilitates the adoption of creative pedagogies and aligns with the needs of modern learners. This study presents a forward-thinking perspective on the future of science assessments in contemporary classrooms, emphasising the utilisation of artificial intelligence (AI) to create tailored and genuine learning experiences that foster motivation among students.

Methodology

The present study employed a qualitative content analysis methodology to investigate the current body of scholarly literature on assessment practises in the field of science education, utilising the AI tool ChatGPT as a lens through which to analyse the data. Content analysis enables a methodical examination and consolidation of textual data in order to discern significant themes, concepts, and ideas (Krippendorff, 2004). To gather relevant articles, major databases including Web of Science, Scopus, Emerald Insights, EBSCO Academic Search Complete, ERIC, and ProQuest were searched in February 2023 using Boolean search strings with keywords such as "science assessment" OR "science evaluation" AND "ChatGPT" OR "artificial intelligence" OR "AI" NOT "STEM" OR "math*" OR "engineer*". Additional papers were identified through manual searches of reference lists.

The inclusion criteria for this study encompassed publications within the timeframe of 2018–2023, written in the English language, and consisting of peer-reviewed articles or conference papers. The primary focus of these publications was on the assessment of scientific knowledge and skills, specifically examining the role of ChatGPT or artificial intelligence (AI) technology in this context. Articles were excluded if they focused exclusively on subjects unrelated to science. The first step involved conducting an initial screening of titles and abstracts. Subsequently, the full texts of selected articles were thoroughly reviewed to establish a corpus of relevant literature for further analysis. The articles underwent analysis in order to identify recurring patterns and significant substantive categories. The conclusions drawn in this study are supported by a thematic analysis of the data, which was conducted using a rigorous qualitative approach. The analysis focused on recent scholarly insights pertaining to the assessment of science using artificial intelligence.

Traditional Assessments in Science

Quansah (2018) and Asad et al. (2021) assert that traditional assessments encompass the customary approaches to evaluating knowledge and skills, typically yielding written artefacts such as quizzes or examinations. These examinations are administered by teachers to assess the extent of students' learning. The commonly employed objective assessment methods encompass multiple-choice tests, true/false tests, and essays (Puthiaparampil & Rahman, 2020; Xu, Kauer, & Tupy, 2016). The main focus of traditional assessments in the field of science is based on behaviourist theories, which prioritise observable and measurable external behaviours (Gavetti et al., 2012; Gupta, 2011). Traditional assessments primarily evaluate learners' ability to recall and reproduce factual knowledge rather than their capacity for deeper comprehension or critical thinking. These objective assessments evaluate students' capacity to retain scientific facts, definitions, or formulas. There is a consensus among scholars that this approach possesses a certain degree of objectivity and facilitates the process of evaluating student performance (Grant & Dweck, 2003; Heffernan, 2022; Quansah, 2018). However, it is important to note that conventional assessment techniques often exhibit a tendency to oversimplify the intricate characteristics inherent in scientific disciplines.

Based on behaviourist principles, the aforementioned assessments often adopted a teacher-centred perspective, wherein the teacher assumed the role of the sole provider of information while the students were anticipated to passively receive and reproduce it (Gavetti et al., 2012; Gupta, 2011; Quansah, 2018). As previously indicated, there is a greater emphasis on rote memorization than active engagement or the practical application of scientific knowledge. During these assessments, students are often required to choose from a limited number of options or provide predetermined responses. There is a significant emphasis placed on providing an accurate response, thereby limiting the learner's opportunity to articulate their ideas or employ higher-order cognitive skills. In addition, it is worth noting that conventional assessment methods have failed to adequately address the scientific inquiry process, which encompasses the formulation of hypotheses, the execution of experiments, and the examination of findings (Pedaste et al., 2015).

In order to assess the extent to which students have retained the material, educators are required to periodically administer assessments, typically following the completion of a unit or chapter. The assessment of teaching strategies' effectiveness is commonly conducted through the utilisation of outcomes, as evidenced by various studies (Heffernan, 2022; Pedaste et al., 2015; Quansah, 2018). Nevertheless, educators are increasingly recognising the necessity for assessments that align more effectively with the goals of science education, thereby shedding light on the limitations of this particular approach. Within this particular context, there has been a notable shift away from behaviourist methodologies towards pedagogical approaches that prioritise student-centred and constructivist principles. This shift aligns with the ongoing advancements in educational practices. The aforementioned shift has prompted a reassessment of traditional evaluation methods, necessitating the adoption of more precise assessment approaches that more effectively capture students' abilities in critical thinking, problem-solving, and the meaningful utilisation of scientific knowledge (Friesen & Scott, 2013; Khalaf & Mohammed Zin, 2018; O'Riordan, Millard, & Schulz, 2021). In the present context, the subsequent section will expound upon the proliferation of authentic assessments within the realm of science. These assessments have demonstrated the capacity to address the limitations inherent in conventional tests while aligning with cognitive learning methodologies.

Authentic Assessments in Science

The conventional methods of assessment exhibit notable constraints, particularly in the context of science education (Ellis et al., 2020; Rudolph, Tan, & Tan, 2023). Consequently, a growing number of educators are opting to utilise authentic assessments, which are in accordance with cognitive theories of learning, as a substitute for traditional behaviourism-based tests (Quansah, 2018). The primary objective of authentic assessments is to assess students' capacity to apply knowledge, engage in critical thinking, and solve problems within real-world contexts (Sokhanvar, Salehi, & Sokhanvar, 2021; Villarroel et al., 2018). This approach facilitates substantial involvement with scientific concepts. In addition, it is worth noting that authentic assessments have a tendency to provide a more precise depiction of students' comprehension. Cognitivism, as a theoretical framework for learning, encompasses the internal cognitive processes involved in the acquisition of knowledge and comprehension ((Mancing & Marston William, 2022; Quansah, 2018).

The argument posits that the process of learning is characterised by active engagement and productivity, wherein students actively manipulate and restructure information in order to construct meaning. The perspective is substantiated by the utilisation of authentic assessments, which assess students' capacity to apply their knowledge in practical contexts and their retention of the subject matter. Authentic assessments that are commonly employed in science classes encompass a range of evaluative methods, such as open-ended questions, lab reports, group projects, case studies, and simulations. In each of these domains, students are required to apply their acquired knowledge, analyse data, make informed judgements, and proficiently convey their findings in order to successfully complete assessments that are designed to emulate authentic scientific methodologies (Quansah, 2018; Wasserstein, Schirm, & Lazar, 2019). By engaging in these activities, students acquire a more profound understanding of scientific principles and develop valuable transferable skills beyond the confines of the classroom.

In the context of open-ended questions, it is the objective of the instructor to elicit thoughtful responses from students rather than mere regurgitation of information (Quansah, 2018). The students are required to demonstrate their comprehension, provide evidence to support their claims, and substantiate their discoveries. This type of assessment has the potential to enhance metacognition and reflection, thereby facilitating learners' evaluation of their cognitive processes, a critical competency in the realm of science education. In the context of laboratory reports and other research endeavours, students are afforded the chance to engage in hands-on experimentation, gather empirical data, and conduct analysis (Hoehn & Lewandowski, 2020; Von Wirth et al., 2019). In addition to assessing students' proficiency in adhering to scientific methodologies, these approaches also evaluate their aptitude for analysing data, drawing conclusions, and effectively communicating their research outcomes. Furthermore, it is crucial to acknowledge that engaging in experiments and analysing the outcomes enhances students' ability to engage in critical thinking, problem-solving, and scientific investigation (Quansah, 2018; Wasserstein et al., 2019).

Moreover, students have the opportunity to employ scientific concepts in practical scenarios within dynamic and authentic environments facilitated by real-world simulations. The assessments in question are designed to simulate real-world scenarios in which students are required to apply their scientific knowledge in order to make informed

decisions and solve complex problems. These scenarios may involve addressing environmental issues, analysing medical cases, or resolving technical challenges (Wu et al., 2013). Simulations offer a secure experimental setting wherein students can observe the outcomes of their decisions and enhance their comprehension through iterative processes of trial and error ((Despeisse, 2018; Merchant et al., 2014). Case studies have been identified as a legitimate assessment approach in the field of scientific education, as indicated by several scholarly sources (Quansah, 2018; Sewagegn & Diale, 2020; Tang et al., 2022). Students are presented with hypothetical or real-world scenarios, which they analyse and resolve based on scientific principles. Case studies replicate real-life situations to enable students to observe the practical application of their scientific knowledge and refine their critical thinking skills (Alam, 2022).

The implementation of authentic assessments in the field of science yields numerous advantages. As students engage in the process of constructing knowledge rather than passively receiving it, it serves to promote active learning (Kalamas Hedden et al., 2017; Strachan & Liyanage, 2015). Authentic assessments have been found to promote the development of a more profound understanding of scientific subjects among students, as they necessitate the engagement of critical thinking, the establishment of connections, and the application of knowledge in practical contexts (Bean & Melzer, 2021; Quansah, 2018). Furthermore, it has been argued that authentic assessments provide students with the opportunity to develop essential skills that go beyond mere knowledge retention. These skills include collaboration, effective communication, data analysis, and problem-solving (Binkley et al., 2012; Darling-Hammond & Adamson, 2014). These assessments are designed to equip students with the necessary skills and knowledge to effectively apply scientific principles in real-world contexts. In such scenarios, students will be required to employ their acquired knowledge and critical thinking to address complex problems.

Need for Another Form of Assessment in Science

The emergence of artificial intelligence (AI) has had a substantial impact on the ongoing process of digitization within society. As per Miller's (2019) research, the concept of artificial intelligence (AI) has been previously characterised as a technological system capable of emulating various human behaviours, such as reasoning, exercising judgement, and exhibiting intentionality. The emergence of machine learning and neural networks has engendered further discourse regarding the definition of artificial intelligence (Zhai, 2023). The aforementioned technology exhibits remarkable capabilities in the automation of tasks, the processing of extensive amounts of data, and the provision of extrapolative insights, thereby leading to a transformative impact on certain facets of individuals' lives (Cooper, 2023). The rapid expansion of artificial intelligence (AI) and the emergence of ChatGPT offer a promising opportunity to enhance science teaching and assessment methodologies. As previously illustrated, conventional assessments often fall short in effectively measuring the criteria of scientific inquiry and, as previously indicated, struggle to actively involve students (Blumenfeld, Kempler, & Krajcik, 2006; Hubbard, 2020). There exists a necessity for the development of a novel form of assessment that incorporates technology while simultaneously upholding the fundamental principles of authentic assessment. This can be achieved through the integration of Artificial Intelligence (AI) and ChatGPT.

Rise of ChatGPT

ChatGPT is an artificial intelligence language model developed by OpenAI. The utilisation of deep learning algorithms enables the generation of responses that closely resemble human-like behaviour, facilitating dynamic and interactive conversations (Cooper, 2023). The technology has garnered significant attention due to its ability to simulate dialogue and provide tailored responses based on contextual cues and user inputs. The entity responsible for the creation of the aforementioned technology is OpenAI. The utilisation of deep learning algorithms facilitates the generation of responses that exhibit human-like qualities and enables active participation in dynamic conversations (Adiwardana et al., 2020; Javaid et al., 2023). The technology has garnered significant attention due to its ability to simulate dialogue and provide tailored responses based on context and user inputs (Baidoo-Anu & Owusu Ansah, 2023; Nuruzzaman & Hussain, 2018).

Kılınç (2023) and Crawford, Cowling, and Allen (2023) have posited that ChatGPT offers distinct advantages within the realm of science education. The system has the capability to function as a chatbot or virtual instructor for students, providing responses to their inquiries and delivering prompt feedback while they engage in inquiry-based learning. ChatGPT provides an interactive platform wherein students can actively participate in the exploration of scientific concepts and apply their acquired knowledge in real-time through the simulation of dialogues (Cooper, 2023; Qureshi, 2023).

Moreover, ChatGPT has the ability to mitigate the limitations of conventional tests and align with the criteria of authentic assessment. Tailoring questions and responses to cater to the specific needs of students can facilitate a more personalised and student-centric instructional approach (Sullivan, Kelly, & McLaughlan, 2023; Tlili et al., 2023). The aforementioned adaptability facilitates the educational needs of students with diverse learning styles and abilities, thereby providing a foundation for differentiated instruction. Cooper (2023) states that ChatGPT possesses the capability to provide expeditious feedback, thereby facilitating students in promptly rectifying any misconceptions and enhancing their understanding in a synchronous manner. The provision of feedback on prom prompts fosters metacognitive awareness, enabling students to engage in reflective thinking and make appropriate modifications to their cognitive processes (Cooper, 2023; Sun & Hoelscher, 2023). Continuous engagement with ChatGPT has the potential to augment students' cognitive processes, such as deep learning, critical thinking, and problem-solving skills (Cooper, 2023; Sun & Hoelscher, 2023).

The utilisation of ChatGPT in genuine assessments presents prospects for diverse and inventive evaluation methodologies beyond the conventional question-and-answer structure (Haleem, Javaid, & Singh, 2022; Nikolic et al., 2023). One potential approach for assessment in the future could involve the implementation of group problem-solving exercises (Lo, 2023).

ChatGPT has the capability to facilitate online group discussions in which students work together to address complex problems or assess real-life scenarios (Kılınç, 2023; Zhu et al., 2023). In addition to assessing the students' understanding of the subject matter, this collaborative evaluation would also measure their abilities in teamwork, communication, and critical thinking. Performance-based assessments are an alternative form of evaluation that employs ChatGPT. In the present context,

it is plausible to request that students exhibit their comprehension and aptitude in these assessments by means of immersive simulations or virtual scenarios. ChatGPT can be employed by students in science classes for various purposes, such as conducting experiments, managing virtual laboratory equipment, and analysing the resulting data (Kılınç, 2023; Sallam, 2023). This approach would provide students with an authentic and immersive context in which to apply their scientific knowledge and skills, facilitating a more accurate assessment of their practical competencies.

ChatGPT can be employed in a variety of ways to assess students in science classrooms. Several instances have been identified in previous studies (Cooper, 2023; Kılınç, 2023): 1) One potential application of ChatGPT is the generation of pre-assessments aimed at evaluating students' existing knowledge on a particular subject matter. The aforementioned data can subsequently be utilised for instructional planning and customization to cater to students who may need supplementary assistance. 2) ChatGPT can be employed to formulate comprehensive assessments aimed at assessing student comprehension and progress subsequent to a designated period of study. This data can subsequently be used to assign grades and offer students feedback, thereby facilitating their academic progress. 3) ChatGPT has the capability to facilitate formative assessments, providing ongoing assistance and feedback throughout the learning journey. Continuous interaction with ChatGPT allows students to receive personalised feedback regarding their comprehension, cognitive processes, and problem-solving approaches (Kasneci et al., 2023; Sun & Hoelscher, 2023). According to the research conducted by Qureshi (2023) and Sallam (2023), students have the opportunity to identify their strengths and limitations by utilising a feedback loop, which enables them to make timely corrections and adjustments. Educators possess the ability to adapt their teaching methods and provide targeted assistance to students based on the insights gained from these interactions ((Byrd & Alexander, 2020).

Uludag (2023) posits that ChatGPT has the potential to be employed in assessments aimed at evaluating an individual's social and emotional intelligence. The incorporation of these competencies into assessments is of utmost importance as their relevance in the field of education is progressively recognised (Crawford et al., 2023; Uludag, 2023). The utilisation of ChatGPT's simulation of social interactions, as outlined in the studies conducted by Bozkurt et al. (2023) and Kılınç (2023), offers students an opportunity to enhance and demonstrate their abilities in empathy, communication, and teamwork. Students have the opportunity to engage in role-playing exercises, engage in discussions about ethical dilemmas, and navigate complex social situations by conversing with ChatGPT. This enables a comprehensive assessment of their social and emotional development (Kasirzadeh & Gabriel, 2023; Kirk et al., 2023).

Framework for using ChatGPT in Science Assessment

Here is a ChatGPT framework for science classroom assessment (see Figure 1).

1. *Determine the Learning Objectives*: Identify the specific information, skills, or proficiencies that correspond with your chosen field of study and the educational prerequisites.

2. *Create Authentic Assessments*: Develop authentic assessment tasks that simulate real-world scenarios, necessitating that students apply their acquired knowledge and skills. The provided examples encompass open-ended inquiries, problem-solving tasks, critical thinking exercises, and case studies. Ensure that the assessments align with the stated learning objectives.
3. *Establish Assessment Criteria*: Establish explicit guidelines and assessment frameworks for the purpose of evaluating student responses. Identify the fundamental components that serve as indicators of proficiency in the evaluated subject matter or abilities.
4. *Make a list of questions or prompts*: Develop a comprehensive compilation of inquiries or stimuli that are designed to elicit purposeful and thoughtful responses from students. These prompts are designed to elicit responses that align with the intended depth and complexity as outlined in the learning objectives. To cater to diverse assessment formats, it is advisable to include a range of question types such as multiple-choice, free-response, and scenario-based questions.
5. *Conduct the Assessment*: Following the prompts, Initiate a dialogue with ChatGPT, emulating an exchange between a student and the AI model. Kindly inquire, solicit elucidations, and furnish instances or evidence to substantiate their responses.
6. *Evaluation of Student Responses*: Evaluate the quality and accuracy of student responses based on the predetermined criteria. When evaluating the quality of information, it is important to take into account various variables, including the accuracy of the information, the depth of knowledge demonstrated, the clarity of explanations provided, and the individual's critical thinking abilities.
7. *Give Feedback*: Based on the students' responses, it is possible to offer constructive feedback by acknowledging their strengths and identifying areas in need of improvement. Please provide additional analysis and expand on your ideas as necessary.

In general, it can be observed that ChatGPT is not serving as a substitute for conventional evaluation methods such as quizzes, examinations, and projects. Nevertheless, ChatGPT can prove advantageous for educators seeking to administer more thorough and captivating evaluations of their pupils (Cotton, Cotton, & Shipway, 2023; Qadir, 2023).

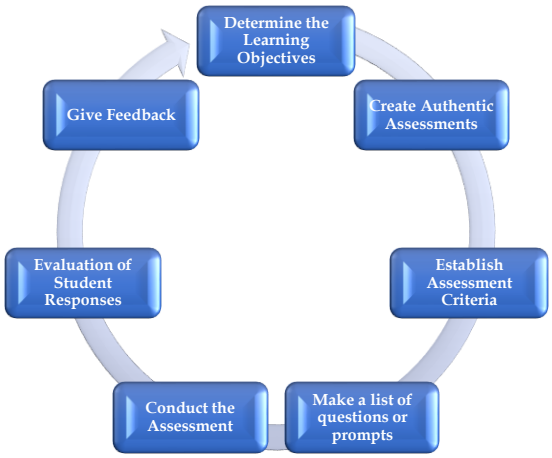

Figure1. *ChatGPT Framework for Science Classroom Assessment*

Implications

The incorporation of ChatGPT into science assessments demonstrates considerable promise in augmenting student engagement, academic achievements, and promoting equitable opportunities for personalised learning encounters. However, in order to fully achieve these advantages, it is necessary to meticulously tackle a variety of sociotechnical factors, ethical quandaries, and practical hurdles associated with implementation. The preservation of student data privacy constitutes a significant apprehension when employing AI-driven assessment technology. To mitigate privacy risks and establish trust, it is imperative to implement robust data governance policies, ensure informed consent procedures, and maintain transparency regarding the utilisation and security of student data by ChatGPT. To prevent instances of unfairness, discrimination, or problematic feedback stemming from limitations in the AI model's training data, it is imperative to incorporate ongoing audits, human oversight mechanisms, and bias mitigation measures. In order to foster comprehensive development, educators must strike a delicate equilibrium between personalised artificial intelligence (AI) assessments and essential interpersonal interactions.

In practical terms, educational institutions may encounter challenges such as insufficient access to devices, issues with network reliability, or a dearth of technical support, all of which impede the seamless integration of ChatGPT, particularly in resource-constrained environments. The incorporation of accessibility and universal design measures will be crucial in ensuring equitable adoption. The swift advancement of AI systems may pose a challenge for educators who are grappling with the ongoing need to adjust their teaching methods and evaluation strategies. The implementation of change management strategies and the provision of ongoing professional development will play a crucial role in providing comprehensive support to teachers. Science educators can effectively manage the risks and optimise the transformative potential of AI in assessments by taking proactive measures to address ethical dilemmas, implementation barriers, and the sociotechnical ecosystem associated with the use of ChatGPT.

However, achieving this goal will necessitate collaborative endeavours among various institutions, meticulous strategizing, and the application of human-centred design principles to develop assessments that are not only tailored to individuals but also considerate of societal implications. As schools transition into the era of AI-enabled learning, it is imperative to conduct additional research on optimal methodologies, maintain ongoing communication, and establish adaptable governance structures.

Recommendations

This paper has examined the potential of conversational AI technologies, such as ChatGPT, to revolutionise science assessment practices. These technologies have the ability to facilitate authentic and personalised learning experiences that are in line with contemporary pedagogical methods. However, the realisation of the complete potential of AI-enabled assessments in a fair and responsible manner necessitates collaborative endeavours spanning institutional boundaries to develop skills, cultivate communities of practitioners, and facilitate ongoing enhancements.

Primarily, it is imperative to allocate significant resources towards teacher training and professional development programmes that prioritise the cultivation of mindsets, skills, and competencies essential for the proficient design, implementation, and evaluation of ChatGPT within science classrooms. Collaboratively developed with assessment experts, researchers, and experienced practitioners, it is imperative that high-quality professional development (PD) programmes strive to cultivate a fundamental understanding of AI among all educators while simultaneously fostering advanced proficiency within leadership ranks across educational institutions. This involves providing teachers with a solid foundation in the ethical aspects of artificial intelligence (AI) as well as practical experience in creating and evaluating ChatGPT assessment interactions. One potential avenue for collaboration between districts and universities is the provision of meticulously designed and easily accessible training modules, workshops, and courses on artificial intelligence (AI) assessment techniques specifically tailored to meet the needs of science teachers.

Furthermore, it is recommended that cross-institutional communities of practice be established at both regional and national levels in order to facilitate collaborative research on the application of artificial intelligence (AI) in the field of science assessments. Consortia have the potential to facilitate the collaboration of various stakeholders, including academia, industry, government agencies, and education non-profits. Through this collaboration, consortia can engage in the examination of evidence-based strategies, the exchange of promising practices, the establishment of consensus standards, and the joint resolution of challenges. Collaborative alliances have the potential to expedite the process of innovation and mitigate redundant endeavours by facilitating the exchange of knowledge. The pilot studies conducted within these communities can be utilised to empirically investigate the various factors that influence the responsible and equitable implementation of AI assessments in diverse classroom settings.

The implementation of targeted funding programmes is crucial in order to address and remove obstacles to educational access, particularly in schools that have limited resources. Grant programmes have the potential to provide financial assistance for the procurement of digital devices, enhancement of network infrastructure, and recruitment of IT support personnel, all of which are essential for the seamless integration of ChatGPT tools. The implementation of cost-sharing models among districts has the potential to optimise economies of scale in the acquisition of novel technologies. In disadvantaged contexts, the provision of access to education can be facilitated through the establishment of public-private partnerships that involve edtech companies, foundations, and social impact funds. These collaborations have the potential to offer innovative financing solutions for addressing the issue of limited access to education.

It is imperative for educational institutions to establish specialised teams comprising experts from various disciplines. These teams should be fully dedicated to managing the process of implementing AI assessments in schools. Their primary responsibilities would include providing instructional design support, offering timely guidance and mentorship to teachers, and collaborating with them to develop learning resources. These teams have the potential to establish communication with district leadership and external partners in order to predict upcoming needs and take proactive measures to create solutions. Ensuring seamless transitions that empower and motivate educators will play a crucial role.

In the end, putting in place participatory design projects with cross-disciplinary teams made up of teachers, developers, assessment experts, and technologists could help people come up with new ideas and speed up the creation of new, morally sound artificial intelligence (AI) testing systems designed for science education. The incorporation of concepts such as AI audits, algorithmic bias testing, and participatory algorithmic governance ought to be integrated within the frameworks of design thinking and scenario simulation methodologies.

Through the implementation of comprehensive and coordinated initiatives that are based on rigorous research and inclusive practices, educators in the field of science can harness the potential of artificial intelligence (AI) to revolutionise the assessment process. This transformation can be achieved by aligning pedagogical objectives with the utilisation of AI while also ensuring that ethical and societal obligations are upheld. This requires going beyond limited technical aspects and embracing a comprehensive approach to address the sociotechnical intricacies of responsibly integrating artificial intelligence. By employing sagacity, anticipation, and collaboration, it is possible to broaden the boundaries of scientific evaluation in order to advance the loftiest principles of enlightened pedagogy.

Acknowledgments The authors express their gratitude to Al Ain University for their collaboration and to the dean of scientific research for his valuable guidance and mentorship.

Availability of data and materials We would have loved to share the data; however, the data is primary, and the authors do not wish to share the data as this may breach participant confidentiality.

Authors' contributions Each of the authors contributed to various aspects of the research process, including research design, study implementation, data collection, data analysis, and manuscript preparation. The authors unanimously endorse the submission of the manuscript for consideration of publication. All of the authors read and approved the final manuscript.

Funding This research received no specific grant from any funding Agency in the public, commercial, or not-for-profit sectors. (There is no funding for this paper).

Informed Consent Statement Not applicable.

Research involving human participants and/or animals This article does not contain any studies with human participants performed by any of the authors.

Conflicts of Interests The authors declare that there are no conflicts of interest pertaining to the publication of this manuscript. Furthermore, the authors have diligently adhered to ethical considerations, encompassing matters such as plagiarism, informed consent, research misconduct, data fabrication or falsification, duplicate publication or submission, and redundancies.

References

Adiwardana, D., Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., & Lu, Y. (2020). Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*. https://doi.org/10.48550/arXiv.2001.09977

Alam, A. (2022). Educational robotics and computer programming in early childhood education: a conceptual framework for assessing elementary school students' computational thinking for designing powerful educational scenarios. In *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICSTSN53084.2022.9761354

Asad, M. M., Khan Soomro, R. B., Shamsy, A., & Churi, P. (2021). Students' satisfaction towards e-assessment for academic achievement in ESL at public schools and colleges. *Education Research International, 2021*, 4576750. https://doi.org/10.1155/2021/4576750

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning (January 25, 2023). *Available at SSRN 4337484*. https://dx.doi.org/10.2139/ssrn.4337484

Bean, J. C., & Melzer, D. (2021). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom*. John Wiley & Sons.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 17-66). Springer Netherlands. https://doi.org/10.1007/978-94-007-2324-5_2

Blumenfeld, P. C., Kempler, T. M., & Krajcik, J. S. (2006). Motivation and Cognitive Engagement in Learning Environments. In R. K. Sawyer (Ed.), *The Cambridge handbook of: The learning sciences* (pp. 475-488). Cambridge University Press. https://doi.org/10.1017/CBO9780511816833.029

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., & Honeychurch, S. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education, 18*(1), 53-130. https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/709

Byrd, D. R., & Alexander, M. (2020). Investigating special education teachers' knowledge and skills: Preparing general teacher preparation for professional development. *Journal of Pedagogical Research, 4*(2), 72-82. https://doi.org/10.33902/JPR.2020059790

Cooper, G. (2023). Examining science education in chatgpt: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444-452. https://doi.org/10.1007/s10956-023-10039-y

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 20*(3), 02. https://doi.org/10.53761/1.20.3.02

Darling-Hammond, L., & Adamson, F. (2014). *Beyond the bubble test: How performance assessments support 21st century learning*. John Wiley & Sons.

Despeisse, M. (2018). Games and simulations in industrial engineering education: a review of the cognitive and affective learning outcomes. In *2018 Winter Simulation Conference (WSC)* (pp. 4046-4057). IEEE. https://doi.org/10.1109/WSC.2018.8632285

Ellis, C., Van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., Rozenberg, P., Newton, P., & Saddiqui, S. (2020). Does authentic assessment assure academic integrity? Evidence from contract cheating data. *Higher Education Research & Development, 39*(3), 454-469. https://doi.org/10.1080/07294360.2019.1680956

Friesen, S., & Scott, D. (2013). Inquiry-based learning: A review of the research literature. *Alberta Ministry of Education, 32*, 1-32. https://www.galileo.org/focus-on-inquiry-lit-review.pdf

Gavetti, G., Greve, H. R., Levinthal, D. A., & Ocasio, W. (2012). The behavioral theory of the firm: Assessment and prospects. *Academy of Management Annals, 6*(1), 1-40. https://doi.org/10.5465/19416520.2012.656841

Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85*(3), 541-553. https://doi.org/10.1037/0022-3514.85.3.541

Gupta, S. (2011). Constructivism as a paradigm for teaching and learning. *International Journal of Physical and Social Sciences, 1*(1), 23-47. https://www.indianjournals.com/ijor.aspx?target=ijor:ijpss&volume=1&issue=1&article=002

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *Benchcouncil Transactions on Benchmarks, Standards and Evaluations, 2*(4), 100089. https://doi.org/10.1016/j.tbench.2023.100089

Heffernan, T. (2022). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education, 47*(1), 144-154. https://doi.org/10.1080/02602938.2021.1888075

Hoehn, J. R., & Lewandowski, H. (2020). Incorporating writing in advanced lab projects: A multiple case-study analysis. *Physical Review Physics Education Research, 16*(2), 020161. https://doi.org/10.1103/PhysRevPhysEducRes.16.020161

Hubbard, D. W. (2020). *The failure of risk management: Why it's broken and how to fix it*. John Wiley & Sons.

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115

Kalamas Hedden, M., Worthy, R., Akins, E., Slinger-Friedman, V., & Paul, R. (2017). Teaching sustainability using an active learning constructivist approach: Discipline-specific case studies in higher education. *Sustainability, 9*(8), 1320. https://doi.org/10.3390/su9081320

Kasirzadeh, A., & Gabriel, I. (2023). In conversation with Artificial Intelligence: aligning language models with human values. *Philosophy & Technology, 36*(2), 1-24. https://doi.org/10.1007/s13347-023-00606-x

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khalaf, B. K., & Mohammed Zin, Z. B. (2018). Traditional and inquiry-based learning pedagogy: A systematic critical review. *International Journal of Instruction, 11*(4), 545-564. https://doi.org/10.12973/iji.2018.11434a

Kılınç, S. (2023). Embracing the Future of Distance Science Education: Opportunities and Challenges of ChatGPT Integration. *Asian Journal of Distance Education, 18*(1), 205-237. https://doi.org/10.5281/zenodo.7857396

Kirk, H. R., Vidgen, B., Röttger, P., & Hale, S. A. (2023). Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalised feedback. *arXiv preprint arXiv:2303.05453*. https://doi.org/10.48550/arXiv.2303.05453

Krippendorff, K. (2004). Measuring the reliability of qualitative text analysis data. *Quality and Quantity, 38*, 787-800. https://doi.org/10.1007/s11135-004-8107-7

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Mancing, H., & Marston William, J. (2022). Cognitivism. In H. Mancing & J. Marston William (Eds.), *Restoring the Human Context to Literary and Performance Studies: Voices in Everything* (pp. 261-280). Springer International Publishing. https://doi.org/10.1007/978-3-030-89078-0_11

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29-40. https://doi.org/10.1016/j.compedu.2013.07.033

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education, 48*(4), 559-614. https://doi.org/10.1080/03043797.2023.2213169

Nuruzzaman, M., & Hussain, O. K. (2018). A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)* (pp. 54-61). IEEE. https://doi.org/10.1109/ICEBE.2018.00019

O'Riordan, T., Millard, D. E., & Schulz, J. (2021). Is critical thinking happening? Testing content analysis schemes applied to MOOC discussion forums. *Computer Applications in Engineering Education, 29*(4), 690-709. https://doi.org/10.1002/cae.22314

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47-61. https://doi.org/10.1016/j.edurev.2015.02.003

Puthiaparampil, T., & Rahman, M. M. (2020). Very short answer questions: a viable alternative to multiple choice questions. *BMC Medical Education, 20*(1), 1-8. https://doi.org/10.1186/s12909-020-02057-w

Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-9). IEEE. https://doi.org/10.1109/EDUCON54358.2023.10125121

Quansah, F. (2018). Traditional or performance assessment: What is the right way to assessing learners. *Research on Humanities and Social Sciences, 8*(1), 21-24. https://www.iiste.org/Journals/index.php/RHSS/article/view/40787

Qureshi, B. (2023). Exploring the use of chatgpt as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. *arXiv preprint arXiv:2304.11214*. https://doi.org/10.48550/arXiv.2304.11214

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342-263. https://doi.org/10.37074/jalt.2023.6.1.9

Sallam, M. (2023). The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. *medRxiv*, 2023.2002. 2019.23286155. https://doi.org/10.1101/2023.02.19.23286155

Sewagegn, A. A., & Diale, B. M. (2020). Authentic assessment as a tool to enhance student learning in a higher education institution: implication for student competency. In *Assessment, Testing, and Measurement Strategies in Global Higher Education* (pp. 256-271). IGI Global. https://doi.org/10.4018/978-1-7998-2314-8.ch013

Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation, 70*, 101030. https://doi.org/10.1016/j.stueduc.2021.101030

Strachan, R., & Liyanage, L. (2015). Active Student Engagement: The Heart of Effective Learning. In P. C. Layne & P. Lake (Eds.), *Global Innovation of Teaching and Learning in Higher Education: Transgressing Boundaries* (pp. 255-274). Springer International Publishing. https://doi.org/10.1007/978-3-319-10482-9_16

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching, 6*(1), 1-10. https://doi.org/10.37074/jalt.2023.6.1.17

Sun, G. H., & Hoelscher, S. H. (2023). The ChatGPT storm and what faculty can do. *Nurse Educator, 48*(3), 119-124. https://doi.org/10.1097/nne.0000000000001390

Tang, Y. M., Chau, K. Y., Kwok, A. P. K., Zhu, T., & Ma, X. (2022). A systematic review of immersive technology applications for medical practice and education-trends, application areas, recipients, teaching contents, evaluation methods, and performance. *Educational Research Review, 35*, 100429. https://doi.org/10.1016/j.edurev.2021.100429

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Uludag, K. (2023). Testing creativity of ChatGPT in psychology: Interview with ChatGPT. *Available at SSRN 4390872*. https://dx.doi.org/10.2139/ssrn.4390872

van Wyngaarden, A., Leech, R., & Coetzee, I. (2019). Challenges nurse educators experience with development of student nurses' clinical reasoning skills. *Nurse Education in Practice, 40*, 102623. https://doi.org/10.1016/j.nepr.2019.102623

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*(5), 840-854. https://doi.org/10.1080/02602938.2017.1412396

Von Wirth, T., Fuenfschilling, L., Frantzeskaki, N., & Coenen, L. (2019). Impacts of urban living labs on sustainability transitions: Mechanisms and strategies for systemic change through experimentation. *European Planning Studies, 27*(2), 229-257. https://doi.org/10.1080/09654313.2018.1504895

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p< 0.05". *The American Statistician, 73*(sup1), 1-19. https://doi.org/10.1080/00031305.2019.1583913

Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., & Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education, 62*, 41-49. https://doi.org/10.1016/j.compedu.2012.10.024

Xu, X., Kauer, S., & Tupy, S. (2016). Multiple-choice questions: Tips for optimizing assessment in-seat and online. *Scholarship of Teaching and Learning in Psychology, 2*(2), 147-158. https://doi.org/10.1037/stl0000062

Zhai, X. (2023). Chatgpt for next generation science learning. *XRDS: Crossroads, The ACM Magazine for Students, 29*(3), 42-46. https://doi.org/10.1145/3589649

Zhu, G., Fan, X., Hou, C., Zhong, T., Seow, P., Shen-Hsing, A. C., Rajalingam, P., Yew, L. K., & Poh, T. L. (2023). Embrace Opportunities and Face Challenges: Using ChatGPT in Undergraduate Students' Collaborative Interdisciplinary Learning. *arXiv preprint arXiv:2305.18616*. https://doi.org/10.48550/arXiv.2305.18616