Computers in Human Behavior: Artificial Humans 2 (2024) 100040

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

A review of assessment for learning with artificial intelligence

Bahar Memarian *, Tenzin Doleck

Simon Fraser University, Faculty of Education, Vancouver, British Columbia, Canada

ARTICLE INFO

Keywords:
Assessment For Learning or AFL
Artificial Intelligence or AI
Education
Review

ABSTRACT

The reformed Assessment For Learning (AFL) practices the design of activities and evaluation and feedback processes that improve student learning. While Artificial Intelligence (AI) has blossomed as a field in education, less work has been done to examine the studies and challenges reported between AFL and AI. We conduct a review of the literature to examine the state of work on AFL and AI in the education literature. A review of articles in Web of Science, SCOPUS, and Google Scholar yielded 35 studies for review. We share the trends in research design, AFL conceptions, and AI challenges in the reviewed studies. We offer the implications of AFL and AI and considerations for future research.

1. Introduction

This work presents a systematic review of the literature to obtain a snapshot of studies, conceptions, and challenges surrounding assessment for learning and artificial intelligence. Our review of 35 studies presents the landscape in which research and conceptions are being carried out toward assessment for learning. The challenge of integrating artificial intelligence with assessment for learning is diverse. The contribution of this work is in shedding light on implications and considerations for future work.

1.1. Overview

In the view of "assessment for learning", assessment is at the core of teaching and learning (Black et al., 1998). This view aims to connect what the teacher expects with what the student needs to demonstrate as evidence of learning and what is being evaluated through communication and the use of formative evaluation criteria. Assessment for learning may or may not provide a summative measure of student learning. But what makes an assessment useful for learning is the promotion of scaffolding and formative feedback for rectification of past errors and future directions for learning growth (Stiggins, 2005).

Assessments have the natural tendency to be more subjective (Sawand et al., 2015). Though standardization and closed-ended assessments such as scoring schemes aim to objectify assessment, they are still heavily dependent on the expectations orchestrated by the assessor and manifested in the learning activities and assessment instructions. Assessments may fail to capture the subjective internal processes that happen within the learner, assessor, and learning experience (McConlogue, 2020). Such internal processes are holistic and may not just affect learning but also target other affective and kinesthetic states of the learners immediately or in the long run within the learning organization (Izci, 2016; Nicol, 2021).

Subjectivity may mean more than just a mere lack of measurement. A search in the education literature can reveal a plethora of assessment tools, most of which are tested and validated through simulated or real settings with samples of variable size (Gikandi et al., 2011; Mandernach, 2015). All such tools are meant to measure, whether it is the student's or teacher's performance, opinions, or other metrics. Indicators such as validity and reliability may give some reassurance of the stable effect of assessment, yet there is almost no guarantee that a new learner, assessor, or learning activity may evoke the assessment measurements.

1.2. AFL underpinnings

Assessment for learning, also known as assessment as learning, formative assessment, learning-oriented assessment, and sustainable assessment (McDowell et al., 2009) is seen as a reformed practice in evaluating student performance and feedback delivery. Influenced by several researchers (e.g., Black & Wiliam, 1998; Carless, 2015; Gibbs, 2006; Gipps & Murphy, 1994; Nicol & Macfarlane-Dick, 2004; Sadler, 1989; Winter 2003), the concept of assessment for learning has evolved (see Table 1) to present a broader and more critical issue of evaluation and feedback delivery that is insightful of student learning gaps and areas that need improvement. In this sense, assessment is not just corrective but also formative of what was learned in the past and what

* Corresponding author.
E-mail address: bmemaria@sfu.ca (B. Memarian).

https://doi.org/10.1016/j.chbah.2023.100040
Received 21 November 2023; Received in revised form 20 December 2023; Accepted 23 December 2023
Available online 29 December 2023

Table 1
Summary of theories of assessment for learning.

Topic	Reference
Various factors including access to education, curriculum, motivation of students, and teacher characteristics all interact with the assessment mode.	Gipps and Murphy (1994)
Assessment practice is changing from assessment of what students know or assessment of learning to the assessment that informs learning or assessment for learning.	Winter (2003)
Propose 11 conditions under which assessment can facilitate student learning: 1. Assessed tasks capture sufficient study time and effort, 2. These tasks distribute student effort evenly across outcomes and weeks, 3. These tasks engage students in productive learning activities, 4. Assessment communicates clear and high expectations to students, 5. Sufficient feedback is provided, both often enough and in enough detail, 6. Feedback focuses on learning rather than on marks or students themselves, 7. The feedback is provided quickly enough to be useful to students, 8. Feedback is linked to the purpose of the assignment and to criteria, 9. Feedback is understandable to students, given their sophistication, 10. Feedback is received by students and attended to, and 11. Feedback is acted upon by students to improve their work or their learning.	Gibbs (2006)
Provide direct authentic opportunities of evaluation for students to experience assessment for learning appropriately.	Sadler (1989)
Formative assessment is an integral component of classroom work and that its enhancement can raise standards of achievement.	Black and Wiliam (1998)
'Feedback' and 'feedforward' should be systematically set in in curriculum practices. Seven principles are proposed by the authors:" 1. Provides opportunities to close the gap between current and desired performance. 2. Helps clarify what good performance is (goals, criteria, or expected standards). 3. Delivers high-quality information to students about their learning. 4. Facilitates the development of self-assessment (reflection) in learning. 5. Encourages teacher and peer dialogue around learning. 6. Encourages positive motivational beliefs and self-esteem. 7. Provides information to teachers that can be used to help shape the teaching." (p. 3)	Nicol and Macfarlane-Dick (2004)
Assessment has been comparatively overlooked in comparison to learning and teaching.	Carless (2015)

needs to change and be learned in the present and future.

1.3. AI underpinnings

Artificial intelligence has been long studied in various disciplines including in education (Memarian & Doleck, 2023). Yet, theories that directly tie technologies such as AI with education, particularly assessment for learning, are not extensively researched. A popular kind of AI, known as generative AI offers (Dwivedi et al., 2023):

- Interpretation and response to questions and prompts
- Translation across languages
- Text generation in different presentation forms
- Analysis and summary of data in various forms such as text

Several methods to classify AI in education have been put forth (Gao, Nagel, & Biedermann, 2019; Van Vaerenbergh & Pérez-Suay, 2022; Zhai et al., 2021), see Table 2 for a list. Yet, when it comes to making connections between learning theories and AI less work has been done. For example, Chen et al. (2020) conducted a review of the literature and found a lack of studies that both employ AI technologies and engage deeply with educational theories.

Table 2
Summary of work around assessment and artificial intelligence.

Topic	Reference
Three paradigms of learning with AI, namely: - Paradigm one is when the learner is the recipient, and the AI directs the cognitive learning, - Paradigm two is when the learner is collaborating with AI for learning and so AI supports cognitive learning, - Paradigm three is when the learner directs the cognitive learning and AI is used to empower the learning.	Ouyang and Jiao (2021)
Three categories of assessment with AI, namely: - "Category 1: AI tools cannot be used; In-person unseen examinations, Class tests, Some online tests, Vivas, Some laboratories and practical, Discussion-based assessments. - Category 2: AI tools can be used in an assistive role; drafting and structure content, supporting the writing process in a limited manner, as a support tutor, supporting a particular process such as testing code or translating content, giving feedback on content, or proofreading content. - Category 3: AI has an integral role; drafting and structuring content, generating ideas, comparing content (AI generated and human generated), creating content in particular styles, producing summaries, analyzing content, reframing content, researching and seeking answers, creating artwork (images, audio and videos), engaging in a conversational discussion, developing code, translating content, generating initial content to be critiqued by students." (p.1)	UCL (2023)

1.4. Gap

Because of the diverse and dynamic uncertainties surrounding assessment measurement, the literature suggests paying more attention to the communication of expectations, assumptions, evidence, and outcomes (Hargreaves, 2005). Assessment for learning focuses on the idea of informing learners about the assessment expectations, collecting evidence concerning the communicated expectations, and making sense of evaluation through qualitatively documented (e.g., through rubric) means. This approach in a way accepts that assessment measurement is susceptible to errors, and instead emphasizes the communication of what happened in the assessment process, whether good or bad. Given the technological advancements such as AI, we are interested in learning more about assessment for learning practices considering AI. To do so we conduct a systematic review of the literature to obtain the current trends and challenges.

1.5. Research questions

We seek to examine the following research questions:

- What are the findings of studies on assessment for learning and artificial intelligence?
- What are the (if any) conceptions of assessment for learning in the reviewed studies?
- What are the (if any) challenges of artificial intelligence reported in the reviewed studies?

2. Methods

We follow a systematic search process and use the thematic coding described below to extract data and chart the papers. Our review aims to find trends and insights on the link between assessment for learning and artificial intelligence in education.

2.1. Search process

We searched the string: "assessment for learning" AND ("artificial intelligence" OR "AI") in the Web of Science (WoS), SCOPUS, and Google Scholar. Inclusion criteria contain studies in English that had a focus on assessment for learning plus artificial intelligence. We removed studies that were not in English and focused on other types of assessments only (e.g., summative assessment) or used AI for assessment of specialized fields such as risk assessment or health assessment. Note that while most of the studies included in the review touched upon assessment for learning or formative assessment, not all may have covered artificial intelligence and instead may just have a citation with text containing artificial intelligence. We thus found our primary inclusion criteria to be focused on studies with a mention of assessment for learning.

The overview of our search process is summarized in the PRISMA chart shown in Fig. 1. A total of 211 (3 from WOS, 8 from SCOPUS, 200 from Google Scholar with 2 duplicates removed) were included for screening. The limited record highlights the little amount of research done at the intersection of assessment for learning and artificial intelligence. Our initial screening examined the title and abstract of 209 studies.

2.2. Protocol

Our process included the following steps:

- Used the input string of search in both SCOPUS and WoS.
- Download the full records of all the papers.
- Identify duplicate studies and remove them.
- Review the title and abstract of each article and decide if relevant or not.
- Download and review the full text of each article and decide if relevant or not.
- Summarize the overview and findings.
- Search for assessment for learning or AFL.
- Extract and summarize the references noted.
- Extract and summarize how AFL is conceived.
- Extract and classify the challenges of AI.

We use grounded theory as our methodological approach and open and axial coding more specifically to identify patterns and trends in the literature. Our goal is thus to code and chart information about assessment for learning and AI challenges in the reviewed studies through both high-level classifications and in-depth analysis.

3. Results

The result section is centered around our three research questions, namely the overview and findings, AFL, and AI issues reported in the reviewed studies.

3.1. Purpose of the reviewed studies and findings

Qualitative studies were the most common type of research design ($N = 28$) followed by mixed-methods ($N = 6$) and quantitative studies ($N = 1$).

3.1.1. Qualitative studies

Baidoo-Anu and Owusu Ansah (2023) review the literature to suggest potential benefits and drawbacks of ChatGPT in promoting pedagogy. The authors find that the use of AI in teaching and learning is more likely. As such the education of both teachers and students on safe and pedagogically sound uses of AI programs such as ChatGPT is needed.

Black et al. (2004) share strategies for teachers to change their practices and students to change their learning behaviors so that the goals of assessment for learning are met appropriately. The authors highlight the need for mechanisms that allow students to compare their performance against themselves, rather than their peers. Giving students comments rather than marks is one way to achieve this.

Dai and Ke (2022) explore the affordances of AI in addressing individual needs and offering active learning experiences in simulation-based learning. Findings offer five areas in which more research with AI-powered simulation-based learning is needed: 1) examine module-based AI and hybrid AI mechanisms for learning; 2) study effects of various designs and scaffoldings with the use of AI-powered virtual agents guided by learning principles and theories; 3)

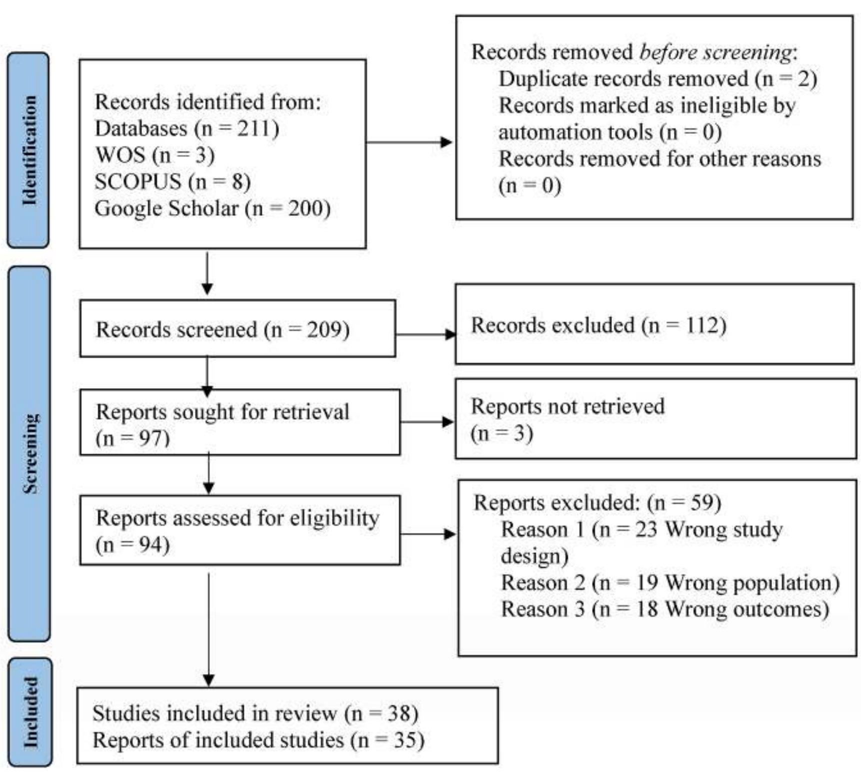

Fig. 1. Prisma chart.

investigate delayed effects of affective states in AI-powered virtual agents; 4) develop machine/deep learning algorithms or techniques for customization, localization, contextualization with small datasets using unbiased and inclusive approaches; and 5) study ethics-related endeavors in AI in simulation-based learning.

DiCerbo (2020) discusses some of the advances in assessment in the digital world and then suggests areas of focus to enable new practices in assessment design to support the assessment of diverse learners. Findings suggest that digital environments can offer more data collection during learning and remove the barrier between assessment and learning.

DiCerbo (2021) explores some of the limitations of intelligent tutoring systems that are adaptive to students' profiles and learning characteristics. The author especially focuses on the scoring algorithms that use a "learned" model of scoring written assignments and shares their issues through a case study of the Khan Academy platform. After several difficulties, primarily due to added complexity and the need for computational power, the Khan Academy platform decided to remove the AI recommendation systems. Instead, there is a mastery system with rules about the number of questions the students need to get right to make a progression.

Dixon et al. (2011) observed teachers teaching a written language unit, particularly the role teachers' beliefs played in the enactment of specific assessments for learning practices. In this study, teachers' beliefs, both those espoused and those held tacitly, were influential concerning their assessment practice. Since beliefs can be private, tacit, unarticulated, and habitual considerable time must be spent during professional development opportunities raising teachers' awareness of their beliefs and how they may influence their practice.

Freeman and Dobbins (2013) reflect on a series of workshops with educators at Birmingham City University and share the Student Enhanced Learning through Effective Feedback (SENLEF) model to support the integration of feedback and feedforward as a social practice in higher education. The authors also outline the principles of the model – reflection, transparency, and developmental dialogue. The authors propose the model as a useful lens through which to view and rethink the current, mechanistic approaches taken in course evaluations.

Fuller et al. (2022) Using feedback from in-person workshops at the 2020 Ottawa Conference as well as open consultation and workshops, the authors propose a consensus framework comprising of: (1) readiness for technology adoption; (2) its application to the assessment lifecycle (supported by illustrative case studies); and (3) processes for evaluation and dissemination of technology-enhanced assessment. One of the key findings of the research was targeted to the application stage and envisioning it into an assessment lifecycle with five foci: (1) Advancing authenticity of assessment; (2) Engaging learners with assessment; (3) Enhancing design and scheduling; (4) Optimizing assessment delivery and recording learner achievement; and (5) Tracking learner progress and faculty activity and thereby supporting longitudinal learning and continuous assessment.

Gamage et al. (2022) examine the practices used in closed and open-book tests and identify the challenges of reformulation into online mode. Findings reveal that online exams may affect students and teachers differently. different tests, open and closed book questions, as well as exam structures (from MCQs to essays), are used where each may test different skills and domains of learning. Students present an interest in alternative assessment methods rather than traditional exams. The authors provide insights on how to rethink assessment strategies during the COVID-19 pandemic.

Hargreaves (2005) shares six groups of definitions for assessment for learning. One definition concerns about monitoring students against target objectives. Another definition explores using assessment to inform the next steps in pedagogy. A third definition suggests teachers giving students feedback for improvement while a fourth definition suggests that teachers learn about the learning mechanisms of students. A fifth definition examines how learners take control of their learning and assessment. In the sixth definition, assessment is recommended to turn into a learning event.

Hargreaves (2007) examines the features of a collaborative assessment for learning. Findings reveal that for an assessment for learning to be valid, its learning outcomes need to be socially appropriate for learners of the twenty-first century. The work shares descriptions of three collaborative assessments for learning currently being practiced (two in the UK and one from the Eastern Caribbean).

Harlen (2005) explores the interplay of summative and assessment learning on one another in high-stakes assessments and on teaching and learning. Authors share that there is value in maintaining the distinction between formative and summative purposes of assessment while seeking synergy about the processes of assessment. The authors provide the analogy of a travel route between two points to present the different purposes assessment can fulfill.

Hawe and Dixon (2017) explore the students' experiences in an environment where students had to take an active role and responsibility for their learning. Using goals can help students understand their learning trajectory and direction. Findings showed that while each AFL strategy provided some effect on student learning, the full impact of AFL specifically on self-regulation was achieved through recursion and iteration.

Hooda et al. (2022) explore the role of qualitative and valid assessment and feedback on students' learning in higher education. A large-scale assessment for learning initiative is implemented in a Swedish municipality. Two founding principles guided this initiative, namely: (1) teaching should be informed by educational research; (2) to be successful teachers' professional development needs to be based on everyday classroom practice. authors share positive results of various assessment and feedback practices that can enhance the student's learning experience and outcomes.

Klenowski (2009) presents an editorial piece on assessment for learning from an Asia-Pacific perspective. An increased interest can be seen in the Asia-Pacific region toward assessment for learning. The authors provide their perspectives on the different terms used in this arena, such as assessment for learning, formative assessment, and practice in a classroom.

Lentz et al. (2021) examine the role of AI in shifting the focus of pedagogy from the assessment of learning to the assessment for learning. Assessment remains a challenge in medical education, and with the advancements in AI, more major shifts are on the horizon. The authors propose the co-production and evaluation of the technology with geographically and culturally diverse stakeholders.

Oladele et al. (2022) provide a review of computer adaptive-based assessment for learning in stem education in Africa. The authors find that using computer-assisted learning and Industry 4.0 technology would further build the need for critical thinking, complex problem solving, and creativity required for Mathematics education.

Renzulli (2021) shares their process of validating an instrument for the assessment of students' executive functions along with the development of a few other tools intended to facilitate students' perceptions and achievement in learning. Their clustering approach revealed interpretable insights into how the test administration decisions were associated with students' performance profiles.

Shute et al. (2017) explore the approach for developing and designing valid assessments and what they call "stealth assessment": reliable and valid inferences about what students know and can do across multiple contexts. The steps in building a stealth assessment in an immersive environment are presented through a worked example.

Swiecki et al. (2022) conducted a review to share some prominent issues of assessment in education. Assessments are often difficult for educators to design and implement, create learning experiences discretized, and not adaptable to individual needs and learning situations. The authors review AI approaches that aim to address these issues and discuss new challenges and considerations.

Taras (2008) provides some of the differences between critical elements of assessment terminology specifically in higher education in the United Kingdom. Findings present that each sector has developed expertise in assessment that has developed from its own specific historical and logistical contexts. This work highlights that there is much to do in the area of assessment to make it more clearly tied to recent developments.

Van der Kleij et al. (2015) address the theoretical differences and similarities among three approaches to formative assessment that are currently most frequently explored in the educational research literature: data-based decision-making (DBDM), AFL, and diagnostic testing (DT). Furthermore, the differences and similarities in the implementation of each approach are examined. Findings show that although theoretical differences exist between DBDM, AFL, and DT, their combined use can develop more informed learning environments.

Wang and Cheng (2021) conduct a scoping review of research studies on artificial intelligence in education (AIED) between 2001 and 2021. The findings of the scoping review suggest three promising AIED research agendas: learning from AI, learning about AI, and learning with AI. Learning from AI: AI serves as the principal means by which students learn. Learning about AI: efforts toward equipping learners to thrive in an AI-saturated future, i.e., Learning about AI (interchangeably, AI education). Learning with AI: using AI tools to improve learning and teaching practice.

Wiliam (2018) provides a commentary on definitions conceived in the literature on assessment for learning. Existing definitions severely limit what counts as an assessment for learning. The findings of the study reconfirm the added benefits of employing assessment for learning or formative assessment strategies as opposed to just the use of summative assessment.

Willis (2011) analyzes examples of assessment for learning practices implemented in the classrooms, while students act as autonomous learners within the learning environment. The work examines the evolution from assessment for learning to assessment as learning from four aspects, namely: participants, test form, multivariate data for process-based measurement, and measurement models for multivariate data.

Yang and Xin (2022) emphasize key attributes of assessment that need to be met. Examples include validity, ethics, and fairness of the measurement. In addition, the authors recommend multidisciplinary corporations in the field, which integrate education, psychology, and information technology into the theories and methods of educational measurement.

Zhai (2021) examines machine learning (ML)–based science assessments and elaborates on how ML innovates assessment practices in science education. The present study makes several noteworthy contributions to research on LOA. Examples include: "ML allows assessments to target complex, diverse, and structural constructs, and thus better approaching the three-dimensional science learning goals of the Next Generation Science Standards (NGSS Lead States, 2013); ML extends the approaches used to eliciting performance and collecting evidence; ML provides a means to better-interpreting observations and using evidence; ML supports immediate and complex decision-making and action-taking." (p. 1).

Zeng et al. (2018) conduct a review of the literature to summarise the state of learning outcomes assessment, namely the history, nature as well as the strategy of developing learning outcomes assessments.

3.1.2. Mixed-methods studies

Bezzina et al. (2021) aim to constructively align gamification and AI to digital assessment for a more personalized and adaptive learning experience for students. The authors explore the impact of AI-powered gamification on students' assessment and learning experience. The findings of the work suggest considerations and challenges for AI-powered gamification in education.

Jonsson et al. (2015) conducted a large-scale implementation of AFL in a Swedish municipality to change how teachers talk about their practices and act on them. The findings of their study revealed that assessment practices are largely teacher-centric, which may both add an individualistic narrative to assessment and a high workload to teachers. In turn, students may take less responsibility for their learning.

Lee (2023) studies students' expressions and peer feedback in a large-scale course by using a provided rubric and AI-enabled evaluation to aid the feedback development and delivery process undergone by the students. The authors find that with AI-enabled evaluation, the provision of feedback can become a sustainable process with students making effective use of feedback for their work and making teachers and students both responsible for teaching and learning.

Pfeiffer et al. (2021) The authors share insights on the body of knowledge and practices in education and AI. The authors share consideration in three areas, namely awareness, impact, and future. Instruments that assess soft skills are often completed by the students themselves, and technology and artificial intelligence may help administer and analyze them.

Westbroek et al. (2020) studied two chemistry and two physics teachers to design and implement two educational frameworks, namely the formative assessment of conceptual understanding, as well as whole task-first differentiating instruction (WTDI). Data analysis shows that all teachers changed their practices permanently and implemented AFL and WTDI.

Wiliam et al. (2004) researched to study the work and outcomes of teachers who took time to develop formative assessment strategies. Key findings include the importance of the teacher-student relationship, viewing AFL as patterns of participation that develop expertise, and learner autonomy as a negotiated learner identity within each classroom context.

3.1.3. Quantitative studies

Shin et al. (2022) utilize a large-scale data set and deep learning framework to optimize the test administration process using clustering approaches. The authors find that by having assessments that are continuous and ubiquitous, students are no longer able to "cram" for an exam. This is because although cramming can provide good short-term recall, it is a poor route to long-term retention and transfer of learning.

3.2. Conceptions of AFL in the reviewed studies

We summarized ways in which AFL is conceived in the reviewed studies. Most of the articles provided a definition of AFL without a technological (e.g., AI) perspective embedded into the definition. A total of 32 studies presented AFL without a technological perspective such as AI and only 3 studies presented AFL with a technological perspective.

3.2.1. AFL definition without a technological perspective

3.2.1.1. Focus on activities. Baidoo-Anu and Owusu Ansah (2023) find activities that provide ongoing feedback to inform teaching and learning are considered formative. Similarly, Bezzina et al. (2021) consider tasks that are more authentic, situated, and experiential to better exhibit assessment for learning. Swiecki et al. (2022) share that a predefined set of items (e.g., problems or questions) needs to be used as part of activities to infer claims about students' proficiency in one or more traits. Harlen (2005) encourages students to participate in activities that provide them with opportunities to exercise control over their learning. Westbroek et al. (2020) find assessment for learning to contribute to the development of metacognitive skills and a feeling of ownership. Shute et al. (2017) also emphasize that assessment for learning requires ongoing collection of data as students interact. Wang and Cheng (2021) find the ongoing assessment activities that allow teachers to monitor student learning on a day-to-day basis and adjust their teaching based on what the students need to be successful can be deemed as appropriate assessments for learning practices. Zhai (2021) also finds collecting feedback to support teachers' instructional decision-making and students' learning to be pedagogically beneficial.

3.2.1.2. Focus on learning insights. Wiliam (2018) considers the use of assessment for learning. should increase learning.

Willis (2011) finds sharing learning goals and criteria with students giving them experience in self-assessment and guiding them with feedback can increase learning. Hawe and Dixon (2017) and Klenowski (2009) suggest using information from dialogue, demonstration, and observation in ways that enhance ongoing learning. Lee (2023) and Taras (2008) suggest that students need to be able to understand the goals they are working towards and can peer and self-assess to improve their work. Lentz et al. (2021) also highlight the need for ongoing constructive feedback and coaching with the goal of improvement. Hargreaves (2007) shares that assessment for learning is appropriate when it has high validity and promotes learning. Renzulli (2021) shares the importance of gathering data, usually from the students themselves, and focuses on students as individuals. Shin et al. (2022) emphasize that formative assessment needs to be inherently diagnostic, informative, and instructional. Wiliam et al. (2004) recognize that increased use of formative assessment leads to higher quality learning, yet they also recognize that the standardized motives and pressure in schools prevent the use of such strategies.

3.2.1.3. Differentiation from summative assessment. Black et al. (2004) find the priority in the design and practice of assessment for learning is to serve the purpose of promoting students' learning. This is different from assessment designed primarily to serve the purposes of accountability, ranking, or certifying competence. Gamage et al. (2022) view assessment for learning as a label for a group of practices that have been shown to help students improve their learning. Similarly, Van der Kleij et al. (2015) find that formative assessment can be seen as an umbrella term that covers various approaches to assessment intended to support learning that have different underlying learning theories. Approaches may involve various principles and intentions that shape assessment uses. Hargreaves (2005) finds it important to distinguish between what characterizes a highly valid summative test and what characterizes a highly valid assessment for learning. Hooda et al. (2022) view assessment for learning to be performed throughout the learning process, whereas summative assessment is frequently performed at the end of all learning activities. Jonsson et al. (2015) recommend sharing criteria with learners, providing constructive feedback, and promoting students' active involvement in assessments.

3.2.1.4. Evolutionary perspectives. Assessment for learning beholds future possibilities that are evolutionary and innovative. DiCerbo (2020) lays out a categorization scheme for thinking about future possibilities of assessment for learning. The authors find there is potential progress being made toward not just "assessment for learning" but learning as assessment. In a similar vein, Fuller et al. (2022) find the pedagogical focus of assessment needs to shift from the sole focus on assessment of learning to a greater representation of assessment for learning and sustainable assessment. Dixon et al. (2011) highlight the many possibilities of formative strategies of assessment that can be considered in future research. Yang and Xin (2022) suggest envisioning processes that can improve understanding of learning based on feedback. The evolutionary possibilities of assessment for learning also hold challenges that need to be resolved. DiCerbo (2021), for example, finds the fairness of AI models is under-studied in assessment-for-learning spaces. Freeman and Dobbins (2013) examine the feedback and feedforward strategies specifically that can be woven together to develop assessment for learning strategies.

3.2.2. AFL definition with a technological perspective

Dai and Ke (2022) share that "(1) sensors (2) multimodal data collection (i.e., body movements, eyes tracking, and/or the combinations), including baseline indicators, (3) machine learning algorithms (i.e., SVM and Neural Networks), and (4) feedback mechanism (i.e., real-time and delayed feedback)" (p. 8) may be needed for the formalization of assessment with AI. Oladele et al. (2022) present a Computer Adaptive Learning (CAL) with an assessment for learning connectivism framework. Pfeiffer et al. (2021) critically analyze the role of adaptive learning methodologies from the teacher's perspective and extend this notion not only to support the learning experience itself but also to assess it, particularly in formative and embedded ways.

3.3. AI issues reported in the reviewed studies

Of the 35 articles on assessment for learning studied, only 17 explored considerations and challenges with the use of AI technology. A total of 18 did not cover AI (Klenowski, 2009; Lee, 2023; Lentz et al., 2021; Oladele et al., 2022; Pfeiffer et al., 2021; Renzulli, 2021; Shin et al., 2022; Shute et al., 2017; Swiecki et al., 2022; Taras, 2008; Van der Kleij et al., 2015; Wang & Cheng, 2021; Westbroek et al., 2020; Wiliam, 2018; Wiliam et al., 2004; Willis, 2011; Yang & Xin, 2022; Zhai, 2021). Among the 17 studies, a diverse set of concerns and challenges were raised as described below.

AI applications such as ChatGPT are undoubtedly changing the landscape of education. Baidoo-Anu and Owusu Ansah (2023) denote that this change is equally likely to cause bad than good. There should as such be more skepticism on the use of AI in education, considering the limited scientific evidence on the impact of AI on the learning of the students and faculty (Bezzina et al., 2021; Fuller et al., 2022).

AI applications enable customized learning and offer personalized learning. Yet, there is still little known on what extent to address individual needs is and is not appropriate (Dai & Ke, 2022). Not to neglect, more personalization and continuous and real-time feedback require greater computational power and thus make AI's feedback process multistep and more complicated (DiCerbo, 2020).

Besides the technological demands, there are social and behavioral demands with the use of AI applications that are still largely unmet. Examples include the lack of trust in the systems, and recognition of existing classroom norms and ethics (DiCerbo, 2021). There is little high-quality, empirical research that looks at the outcomes of such technology on learners and faculty (Fuller et al., 2022). From a more pedagogical perspective, AI applications require an understanding of the learning situations and their needs for individuals with a diverse range of abilities, backgrounds, and needs (Hooda et al., 2022).

Feedback resulting from the assessment for learning systems equipped with AI thus needs to be sustainable (Lee, 2023) and acknowledge the uncertainty of predicting the future (Lentz et al., 2021) and degree of variations in learning settings among students (Pfeiffer et al., 2021); yet design future and possibly life-long learning with a humane perspective in mind.

If AI applications are used as tools to measure and synthesize large volumes of self-reported data, they may largely remain biased, mainly because they do not address and do not aim to normalize the subjective nature of data (Renzulli, 2021). This can be particularly challenging where there is no norm or normalization concept and instead, an understanding of diversities is needed.

With that being said, AI is racing to be used in all facets of the educational system and may thus impact all levels of education (Shin et al., 2022). In assessment for learning particularly, we find the artifacts and processes to be dramatically changing from traditional assessment (Swiecki et al., 2022). The current state of research in AI education can therefore have several limitations. Examples include: 1) disconnect between AIED and AI technology; 2) disparity of AIED in educational settings; 3) underrepresentation of AIED in some contexts (e.g., global south); 4) imbalanced disciplinary development; 5) learning from and with AI is less explored in the literature; 6) disconnect between existing educational approaches and technology; and 7) ethics, bias, privacy issues (Wang & Cheng, 2021). Research into the scientific inner workings of learning such as neuroscience may help shed light on some of the limitations (Zeng et al., 2018). Until neuroscientific breakthroughs are reached, however, the noted challenges may be exacerbated in the event of the monopolization of certain algorithms and techniques such as machine learning (Zhai, 2021), making set ways of pedagogy more probable (Yang & Xin, 2022).

4. Discussion

The discussion section shares an overall summary of findings, implications, and considerations for future studies.

4.1. Summary

An overview of our systematic search showed Qualitative studies were the most common type of research design ($N = 28$) followed by mixed-methods ($N = 6$) and quantitative studies ($N = 1$). Often the studies examined the notion or role of assessment for learning through reviews or small-scale studies of learners. The little quantitative work on assessment for learning may signify the gap between AFL/AI theory and formative notions of meaning-making and feedback. Most of the articles defined AFL without a technological (e.g., AI) perspective embedded into the definition. A total of 32 studies presented AFL without a technological perspective such as AI and only 3 studies presented AFL with a technological perspective. Key perspectives in which AI is explored are focus on activities, focus on learning insights, differentiation from summative assessment, and evolutionary perspectives. Of the 35 studies, only 17 explored the challenges of assessment for learning. The studies and notions of assessment for learning tended to be more or less scoped around qualitative studies and emphasize formative feedback. On the other hand, the challenges of technology such as AI integration with assessment for learning were broad and diverse.

4.2. Implications and future considerations of AFL challenges reported in the reviewed studies

Several factors can cause challenges in technology-enhanced assessment. Examples include lack of human interaction, limited understanding, bias in training data, lack of creativity, dependency on data, lack of contextual understanding, limited ability to personalize instruction, and privacy (Baidoo-Anu & Owusu Ansah, 2023). AI programs are susceptible to glitches and creating fake or misinformation. There is a need to understand when ChatGPT is more prone to create invalid information and instruct AI to self-correct its learning.

An empirical study of AI-powered gamification can become too personalized and difficult to compare unless conducted on a wide scale (Bezzina et al., 2021). Yet testing without large-scale implementation is not possible, leading to scall-scale and fragment analyses.

Renovating old and established grading mechanisms and replacing them with comments and formative ways of assessment can become difficult especially if policies are not in place (Black et al., 2004). Standardized and numerical assessments may be perceived to be more objective and rigorous. However, they are of little formative value, if the goal is to make students aware of their learning.

Dai and Ke (2022) share that the fidelity and authenticity of human-AI conversations still need more work. The authors share trends in the use of assessment practices with humans and AI agents. Their findings suggest that multimodal computing can be used for assessment and feedback. Yet, it is still unknown what measures and mechanisms of bodily conditions and gestures should be considered formative of learning in situated and simulation-based environments.

Existing assessments “(1) do not help inform classroom instruction, (2) do not make accurate inferences about diverse learners, and (3) the things they ask learners to do are far removed from the real-life applications of knowledge and skill we desire them to be able to master" (p. 4) (DiCerbo, 2020). While digital environments may enrich data collected from students during learning, they pose challenges and complexities in terms of how the collected data should be discretized and taken into consideration.

A challenge of educational systems that seek practice to mastery, such as that of Khan Academy is in the possible conflict it may have with percentage grading systems and pacing that is fixed (DiCerbo, 2021). Advanced intelligent tutoring systems imply that the technology can become so demanding that causes the removal of the system and the best solution. This is seen in the case of Khan Academy, where the mapping and recommendation system with new content, fitting with the classroom system, and distrust all led to voting out the AI recommendation system.

Staying true to the prespecified criteria and not getting swayed by additional errors or differences that come across as wrong to our beliefs can be challenging in formative assessment (Dixon et al., 2011). Unpicking teacher beliefs may be difficult because as teachers unlearn new beliefs, they may come to pick up new ones.

A challenge of educational systems is that they may promote tokenistic methods that lead to students' passive, rather than active, participation (Freeman & Dobbins, 2013). It is often more difficult than simply having evaluation informed and owned jointly by students and educators.

More work is needed for a truly transformative technology-enhanced assessment. For example, the assessment lifecycle should not just be considered, but also the needs of users in the pursuit of authentic assessment using technology (Fuller et al., 2022). EDUCAUSE is a nonprofit association whose mission is to advance higher education using information technology. Educause proposes three models of ‘restore,’ ‘evolve,’ and ‘transform’ to investigate how digital technologies may continue to shape education, and assessment (Educause, 2023). It is important though to know what aspects of pedagogy are worth restoring, what aspects have good enough foundations to evolve, and what aspects are better off to be transformed.

Not all practices of "learning how to learn" may be transferable and effective across different contexts (Gamage et al., 2022). Thus, contextualizing such generic and subject-specific features is still underdeveloped. In addition, student motivation and attitudes toward formative feedback may vary on a contextual and even daily basis. The many forms and mediums of formative feedback for student learning may be needed.

When placing one educational practice at the forefront of pedagogy (e.g., assessment for learning) other discourses of learning may be squeezed into the background. A healthy balance and use of discourses are thus needed (Hargreaves, 2005). More importantly, the relationship between the different discourses of learning (e.g., self-regulation and AFL) from a practical and not philosophical/theoretical lens is needed.

Teachers may find it difficult to change students' learning behaviors (Hargreaves, 2007). For example, students may be motivated to compete rather than collaborate. This can be particularly problematic in classrooms where competition and hierarchical relationships are entrenched.

A great deal of support and courage are needed from the teachers to change their practices from being test-oriented to being learning-oriented (Harlen, 2005). Higher education institutions may need to be more adaptive and supportive of failures that may happen during assessments designed by teachers with AI. Similarly, students may at times fail to learn with AI. Such failures should not promote a ban culture but instead, allow investigating what makes failures detrimental and/or productive to learning.

Empirical and simulation-based studies may contain inaccuracies or oversimplifications of what may happen in the real world (Hawe & Dixon, 2017). While empirical studies attempt to control situations as much as possible, they may lack transferability and generalizability in the context of student learning.

Often features of assessment like validity and reliability are completely missing in formative assessments (Hooda et al., 2022). As the name suggests, formative assessments are less objectively measurable and are subjectively perceived and communicated by individuals. A point of concern and challenge is thus quantifying what and how to characterize validity and reliability in formative feedback.

Studies on formative assessment may further be biased by the teachers' and students' perceptions and responses who participated in such studies (Jonsson et al., 2015). There is a challenge to decipher what remains generalizable and what remains personal to students.

Auto-piloting assessment and feedback may give an inhumane feel to the learners (Klenowski, 2009). There is also still little known on the impact and roles of interactions between instructors and student-instructors when removed or substituted in the learning environments (Lee, 2023). Learners may as a result withdraw from the technology and resort back to human feedback providers, even if the technology shows itself to be fairer and more objective.

Integration of AI as a collaborator may significantly impact specialized fields such as medical education (Lentz et al., 2021). This can be problematic if the system such as the healthcare model is trapped in its past structures and traditions, despite the evolution of the technology it is embedded with.

Much of the research on computer-assisted learning in mathematics classrooms is done at the basic and secondary levels (Oladele et al., 2022). More work is thus needed to examine the effect of AFL at the higher education level and across different disciplines.

The role of the teachers in the self-directed learning of students while using adaptive educational applications needs to be further investigated (Pfeiffer et al., 2021). This may lead to new roles for the teacher, morphing from a facilitator to a moderator of personalized learning technology.

Gifted talents may not be equally studied in education. For example, there are underrepresentations of low-income and minority students as well as students who have been labeled twice exceptional (extremely high ability while simultaneously being challenged with learning disabilities) (Renzulli, 2021).

Numerous time series clustering algorithms in the literature may exist (Shin et al., 2022). Such algorithms may compete for business-driven key performance indicators as opposed to learning outcomes. Defining adequate learning outcomes and retention from formative feedback may need to be considered in algorithms.

Future assessments need to also make sense of different learning progressions (types of learning and intelligence) of students (Shute et al., 2017). In short, students may start (have different prior knowledge) and progress differently throughout learning activities. AI and the teacher may thus need to be able to guide and facilitate the student's learning and confusion appropriately. AFL checks could support students in assessing themselves andbetter deciding which learning route suits students (Westbroek et al., 2020).

There are challenges with the implementation of assessment for learning and further considerations for contexts of application (Wiliam, 2018). A challenge, for example, is the reformulation of long-standing discretized assessment practices to ones that are more continuous and formative (Swiecki et al., 2022). An analysis of the processes of assessment across courses, disciplines, and sectors may thus be required (Taras, 2008).

An integrated formative assessment approach may have demands such as professional development that need to be met. More work needs to be done on the integration of approaches, such as the three mentioned, for effective outcomes (Van der Kleij et al., 2015). There needs to be more understanding of why, how, and when assessment should be used and by whom. This may become more challenging in the presence of both AI and human teachers.

Learning from AI may lead to bionic humans, whereas learning about AI may lead to humans with dual expertise in AI, and learning with AI may lead to intelligent educational robots (Wang & Cheng, 2021). Strengthening research in each of the fields in isolation may lead to different trajectories and qualities of human life which can in turn pose both benefits and challenges.

While the benefits of assessment for learning or formative assessment strategies are recognized, less is known about how to support teachers in developing their practice (Wiliam et al., 2004). An argument for formative assessment being challenging is that it can provide feedback but not a measure of how the student is performing. While efforts such as categorized rubrics may come to bridge summative and formative assessments, they may still seem more subjective than standardized summative assessments. An implication of formative assessment, which is mostly an added advantage, is that assessment may become more criterion-based and less comparative/competitive.

The anticipated outcome of increased learner autonomy has been elusive rather than conclusive (Willis, 2011). Important factors such as validity, fairness, and ethics issues in formative assessment need to be considered (Yang & Xin, 2022). This can be particularly challenging given the gaps that exist between assessment practices and educational theories (Zhai, 2021).

4.3. Implications and future considerations of AI challenges reported in the reviewed studies

Of the 35 articles on assessment for learning studied, only 17 explored considerations and challenges with the use of AI technology. Baidoo-Anu and Owusu Ansah (2023) and Hooda et al. (2022) are concerned that the ability of AI applications such as ChatGPT to be used in education has caused mixed opinions about it. Educators feel it can revolutionize the existing educational praxis and this may turn good or bad. How AI may thus revolutionize education need to be forecasted and accounted for as much as possible and interventions need to be in place to address the aftermath of bad outcomes caused by AI applications.

Bezzina et al. (2021) highlight that various AI education or AIEd systems have been developed over the years, yet limited scientific evidence of the impact on the quality of learning from such systems is available. This could be partly because key performance indicators or KPIs of technologies and educational systems can be diverse. The KPIs used by industries developing educational systems may need to be less business-driven or also augment learning KPIs, in turn making an understanding of the educational impact of such systems simplified or directly measurable.

Dai and Ke (2022) share that intelligent agents play an important role in addressing individual needs in simulation-based learning as a form of AI. Such needs may be directly related or counter to learning, which makes the differentiation of the needs by AI applications difficult. AI applications may need instructions on how to differentiate between the necessary and unnecessary needs for learning.

DiCerbo (2020) is concerned that AI applications may support immediate processing and automated inferences of student performances, but by doing so they become a multistep and complicated process. There is a challenge of generating superfluous data from learning by AI applications. And so, AI applications may more broadly need instructions on what to consider essential and insightful to learning.

DiCerbo (2021) finds issues remaining in the integration of intelligent tutoring systems and automated scoring in three areas, namely: lack of trust in the systems, lack of consideration of the existing classroom norms, roles, relationships, and responsibilities, and issues of fairness. AI application efforts may need to be placed in areas that add transparency and expansibility, particularly in ways that are less formulaic and more humanly argumentative.

Fuller et al. (2022) share that there is little high-quality, empirical research that studies the outcomes of such AI technology on learners and faculty. What makes the study complicated is AI is largely unique in its adaptive and personalized capabilities. Yet such diverse capabilities can make assessing the impact of AI technology difficult as outcomes may be so-called "apples and oranges".

Lee (2023) shares that using assessment for learning with AI technology requires sustainable feedback mechanisms which may be missing from studies that take a snapshot and limited examination of such technologies. Research on this topic may require more thorough and longitudinal expressions and examinations of feedback.

Lentz et al. (2021) express that the uncertainty of the future cannot be understood, but AI programs can be envisioned and bring the future lovingly into being. This view highlights the importance of a humane and ethical future predicted instead of the capabilities that are offered by AI.

Pfeiffer et al. (2021) explore the capabilities of AI applications to accommodate for variations of each student. A point of challenge becomes the unfamiliarity of AI applications to know where to stop in terms of meeting the unique needs of each student. AI applications may thus need to know how to walk the fine line between what is necessary or not when it comes to meeting the needs of students.

Renzulli (2021) proposes assessment soft skills can be more readily assessed and summarized with the use of AI applications. While such mechanisms offer a picture of what students and other user groups in educational settings experience, they also run the risk of portraying self-reported data.

Shin et al. (2022) denote that AI applications are prone to affecting all levels of education. Such effect may not take place concurrently and as a result, different aspects of the educational system will witness different levels and times of upgrades by such AI technologies.

Swiecki et al. (2022) are concerned that AI-enabled assessment uses dramatically different artifacts and processes from traditional assessments and this transformation may be so rapid that students find the newly implemented processes and artifacts foreign to learning.

Wang and Cheng (2021) share several limitations with the use of AI applications. Examples include: 1) disconnect between AIED and AI technology; 2) disparity of AIED in educational settings; 3) underrepresentation of AIED in some contexts (e.g., global south); 4) imbalanced disciplinary development; 5) learning from and with AI is less explored in the literature; 6) disconnect between existing educational approaches and technology; and 7) ethics, bias, privacy issues. AIED research may need a push to catch up with the technological advancements in AI technology. What makes this process difficult is in determining whether the push is in the right direction and does not lead to obsolete pedagogical approaches over time (Yang & Xin, 2022).

Zhai (2021) examines that machine learning may be a super "bridge" to connect the learning goals and educational decision-making which potentially could attend to the goal of redefining assessment practices. Yet, efforts need to be made to avoid the monopolization of a set approaches in educational settings. This may require making educational industries more responsible and transparent about the impact of their AI-enabled assessment and overall technologies on learning.

4.4. Further considerations

Our review of the literature with search string particularly focused on assessment for learning yielded 35 studies. The nature of studies was most often qualitative and least often qualitative and explored the notion of AFL through focus on activities, focus on learning insights, differentiation from summative assessment, and evolutionary perspectives. To our surprise, the reviewed studies did not necessarily elaborate on established assessment for learning theories in the literature (e.g., Black & Wiliam, 1998; Carless, 2015; Gibbs, 2006; Gipps & Murphy, 1994; Nicol & Macfarlane-Dick, 2004), see Table 1 for overview of each. The challenges noted with AFL in light of AI expanded and were shown to be more diverse, complex, and broader than the ones outlined in AFL literature in education. We thus find there may be a disconnect between the theory and application of assessment for learning in a purely educational versus educational and technological context. Technology such as AI may streamline more real-time and continuous measurement of student data. Yet, an issue remaining is the appropriate collection of data and understanding of what truly qualifies as a learning outcomes assessment.

Future work can benefit from the study of assessment for learning practices that tie technology with theory. For example:

1. Assessed tasks capture sufficient study time and effort: Use multimodal data to capture student time spent in learning management engines, and questions asked with chatbots in engines.
2. These tasks distribute student effort evenly across outcomes and weeks: Use smart assistants in learning management engines that offer student support on how to use their time and resources and learning background efficiently.
3. These tasks engage students in productive learning activities: Present learning activities that match with the learning habits and attributes of learners using screening of student behavior in smart ways.
4. Assessment communicates clear and high expectations to students: Help students understand the breadth and depth of space of quality in learning and performance and use smart approaches to motivate students to travel in an upward trajectory.
5. Sufficient feedback is provided, both often enough and in enough detail: Ensure the feedback is provided in appropriate chunks, sequences, and instances of time.
6. Feedback focuses on learning rather than on marks or students themselves: Enable students to collaborate with a smart artificial tutor/mentor who does not have summative but formative motives for communicating and interacting with the learner.
7. The feedback is provided quickly enough to be useful to students: Understand the interaction of feedback type and time to ensure data is communicated with the students at the right times that lead to maximum learning gain.
8. Feedback is linked to the purpose of the assignment and to criteria: Make the assessment lifecycle apparent to the learner and smart algorithms, helping the AI to be more explainable to the human users.
9. Feedback is understandable to students, given their sophistication: Make sure the level and form of feedback is best delivered and is easily digestible by students of different backgrounds and language skills.
10. Feedback is received by students and attended to: Measure and track the degree of student involvement with the feedback delivered.
11. Feedback is acted upon by students to improve their work or their learning: Elicit how student take and make sense of formative feedback to improve their learning outcomes.

4.5. Limitations of our review

We acknowledge that this work comes with limitations. Our string of searches focused solely on assessment for learning and AI, so alternative and related terms such as formative assessment and Machine Learning were excluded from our study. Further, our review examined the landscape of assessment for learning in studies conducted in English. Other regions may use other terms for assessment for learning or provide new perspectives that are not included in this review. However, the review of references across the 35 reviewed studies provided greater insight into the landscape of assessment for learning in the literature.

5. Concluding remarks

This systematic review aims to provide insight into the studies, findings, conceptions, and challenges in assessment for learning in relation with artificial intelligence. The contribution of this work is in informing challenges, implications, and considerations of assessment for learning and artificial intelligence for future research.

Data availability statements

Data sharing does not apply to this article as no datasets were generated or analyzed during the current study.

CRediT authorship contribution statement

Bahar Memarian: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tenzin Doleck:** Writing – review & editing, Supervision, Software, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Canada Research Chair Program and the Canada Foundation for Innovation.

References

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*. Available at: SSRN 4337484 (2023).
Bezzina, S., Pfeiffer, A., & Dingli, A. (2021). *AI-enabled gamification for learning and assessment*.
Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(2), 8–21.
Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.
Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education, 69*(6), 963–976. https://doi.org/10.1007/s10734-014-9816-z
Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Article 100002.
Dai, C. P., & Ke, F. (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers and Education: Artificial Intelligence*, Article 100087.
DiCerbo, K. (2020). Assessment for learning with diverse learners in a digital world. *Educational Measurement: Issues and Practice, 29*(3), 90–93.
DiCerbo, K. (2021). Why not Go all-in with artificial intelligence? *International Conference on Human-Computer Interaction*, 361–369.
Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting assessment for learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice, 18*(4), 365–379.
Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., … Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642.
Educause. (2023). *Educause*. https://www.educause.edu/.
Freeman, R., & Dobbins, K. (2013). Are we serious about enhancing courses? Using the principles of assessment for learning to enhance course evaluation. *Assessment & Evaluation in Higher Education, 32*(2), 142–151.
Fuller, R., Goddard, V. C., Nadarajah, V. D., Treasure-Jones, T., Yeates, P., Scott, K., … Pyorala, E. (2022). Technology enhanced assessment: Ottawa consensus statement and recommendations. *Medical Teacher, 44*(8), 836–850.
Gamage, S. H. P. W., Ayres, J. R., & Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. *International Journal of STEM Education, 9*(1). https://doi.org/10.1186/s40594-021-00323-x
Gao, P. P., Nagel, A., & Biedermann, H. (2019). Categorization of educational technologies as related to pedagogical practices. *Pedagogy in basic and higher education-current developments and challenges*.
Gibbs, G. (2006). How assessment frames student learning. In C. Bryan, & K. Clegg (Eds.), *Innovative assessment in higher education*. Routledge. https://doi.org/10.4324/9780203969670, 43–56).
Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351.
Gipps, C. V., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Open University Press.
Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213–224.
Hargreaves, E. (2007). The validity of collaborative assessment for learning. *Assessment in Education, 14*(2), 185–199.
Harlen, W. (2005). Teachers' summative practices and assessment for learning–tensions and synergies. *Curriculum Journal, 16*(2), 207–223.
Hawe, E., & Dixon, H. (2017). Assessment for learning: A catalyst for student self-regulation. *Assessment & Evaluation in Higher Education, 42*(8), 1181–1192.
Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering, 2022*. https://doi.org/10.1155/2022/5215722WE-ScienceCitationIndexExpanded(SCI-EXPANDED
Izci, K. (2016). Internal and External factors affecting teachers' adoption of formative. *Assessment to Support Learning, 10*(8), 2774–2781.
Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of assessment for learning in Sweden. *Assessment in Education: Principles, Policy & Practice, 22*(1), 104–121.
Klenowski, V. (2009). *Assessment for learning revisited: An Asia-Pacific perspective*.
Lee, A. V. Y. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies In Educational Evaluation, 77*, Article 101250.
Lentz, A., Siy, J. O., & Carraccio, C. (2021). AI-ssessment: Towards assessment as a sociotechnical system for learning. *Academic Medicine, 96*(75), S87–S88.
Mandernach, B. J. (2015). Assessment of student engagement in higher education: A synthesis of literature and assessment tools. *International Journal of Learning, Teaching and Educational Research, 12*(2), 1–14.
McConlogue, T. (2020). *Assessment and feedback in higher education: A guide for teachers*. UCL Press.
McDowell, L., Sambell, K., & Davison, G. (2009). *Assessment for learning: A brief history and review of terminology*.
Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence, and Higher Education: A Systematic Review. *Computers & Education: Artificial Intelligence, 5*, 100152. https://doi.org/10.1016/j.caeai.2023.100152
Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education, 46*(5), 756–778.
Nicol, D., & Macfarlane-Dick, D. (2004). Rethinking formative assessment in HE: A theoretical model and seven principles of good feedback practice. In B. C. Juwah, D. Macfarlane-Dick, B. Matthew, D. Nicol, & D. Smith (Eds.), *Enhancing student learning though effective formative feedback*. The Higher Education Academy.
Oladele, J. I., Ndlovu, M., & Ayanwale, M. A. (2022). Computer adaptive-based learning and assessment for enhancing STEM education in Africa: A fourth industrial revolution possibility. *Mathematics Education in Africa: The Fourth Industrial Revolution*, 131–144.
Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, Article 100020.
Pfeiffer, A., Bezzina, S., Dingli, A., Wernbacher, T., Denk, N., & Fleischhacker, M. (2021). Adaptive LEARNING and assessment: From the TEACHERS'PERSPECTIVE. *INTED2021 Proceedings*, 375–379.
Renzulli, J. S. (2021). Assessment for learning: The missing element for identifying high potential in low income and minority groups. *Gifted Education International, 3*(2), 199–208.
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.
Sawand, F. A., Chandio, B. A., Bilal, M., Rasheed, M. R., Raza, M. A., & Ahmad, N. (2015). Quality assessment in higher education. *International Letters of Social and Humanistic Sciences, 50*, 162–171.
Shin, J. N., Chen, F., Lu, C., & Bulut, O. (2022). Analyzing students' performance in computerized formative assessments to optimize teachers' test administration decisions using deep learning frameworks. *JOURNAL OF COMPUTERS IN EDUCATION, 9*(1), 71–91. https://doi.org/10.1007/s40692-021-00196-7
Shute, V., Rahimi, S., & Emihovich, B. (2017). Assessment for learning in immersive environments. *Virtual, Augmented, and Mixed Realities in Education*, 71–87.
Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324–328. https://doi.org/10.1177/003172170508700414
Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., … Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3*, Article 100075.
Taras, M. (2008). Assessment for learning: Sectarian divisions of terminology and concepts. *Journal of Further and Higher Education, 32*(4), 389–397.
UCL. (2023). *Using AI tools in assessment*. https://www.ucl.ac.uk/teaching-learning/generative-ai-hub/using-ai-tools-assessment# AI tools cannot be used in this assessment.
Van Vaerenbergh, S., & Pérez-Suay, A. (2022). A classification of artificial intelligence systems for mathematics education. In *Mathematics education in the age of artificial intelligence: How artificial intelligence can serve mathematical human learning*. Cham: Springer International Publishing.
Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. (2015). Integrating data-based decision making, assessment for learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice, 22*(3), 324–343.
Wang, T., & Cheng, E. C. (2021). Towards a tripartite research agenda: A scoping review of artificial intelligence in education research. *International Conference on Artificial Intelligence in Education Technology*, 3–24.
Westbroek, H. B., van Rens, L., van den Berg, E., & Janssen, F. (2020). A practical approach to assessment for learning and differentiated instruction. *International Journal of Science Education, 42*(6), 955–976.
Wiliam, D. (2018). *Assessment for learning: Meeting the challenge of implementation*.
Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*(1), 49–65.
Willis, J. (2011). Affiliation, autonomy and assessment for learning. *Assessment in Education: Principles, Policy & Practice, 18*(4), 399–415.

Winter, J. (2003). *The changing prepositions of assessment practice: Assessment of, for and as learning*.

Yang, L. P., & Xin, T. (2022). Changing educational assessments in the post-COVID-19 era: From assessment of learning (AoL) to assessment as learning (AaL). *Educational Measurement: Issues and Practice, 41*(1), 54–60.

Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning—a critical review of literature. *Educational Assessment, Evaluation and Accountability, 30*, 211–250.

Zhai, X. (2021). Practices and theories: How can machine learning assist in innovative assessment practices in science education. *Journal of Science Education and Technology, 30*(2), 139–149.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., … Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity, 1–18*.