

Artificial Intelligence in the Service of Learning Assessment in Science: Contributions and Challenges Through a Literature Review (2020-2025)

Chaymae Bouzidi¹, Ibtissam Benrriah², Abdelali Kaaouachi³

Science Step Journal / SSJ

2026/Volume 4 - Issue 13

To cite this article: Bouzidi, C., Benrriah, I., & Kaaouachi, A. (2026). Artificial Intelligence in the Service of Learning Assessment in Science: Contributions and Challenges Through a Literature Review (2020-2025). Science Step Journal, 4(13). ISSN: 3009-500X. <https://doi.org/10.5281/zenodo.22113391>

Abstract

This study aims to examine the use of Artificial Intelligence (AI) in science learning assessment by identifying the tools and techniques currently employed, the types of learning that can be effectively assessed, and the advantages and limitations associated with AI-based assessment practices in science education. A systematic literature review was conducted, analyzing 20 recent studies published between 2020 and 2025, focusing on AI applications in science learning assessment contexts. The findings reveal that AI effectively assesses various types of scientific learning, including conceptual understanding, practical skills, scientific argumentation, and problem-solving. Compared to traditional methods, AI offers significant advantages such as real-time feedback, personalized learning experiences, and reduced teacher workload. However, persistent challenges remain, including algorithmic biases, complexity of student responses, and ethical concerns. The thoughtful integration of AI into science learning assessment has the potential to transform education significantly. Nevertheless, careful attention must be paid to ensure its ethical and equitable use in educational settings.

Keywords: Artificial intelligence (AI), Learning Assessment, Science Learning.

¹ PhD Candidate, Mohammed First University, Faculty of Letters and Human Sciences, Oujda, Morocco
c.bouzidi-d22@ump.ac.ma

² PhD Candidate, Mohammed First University, Faculty of Letters and Human Sciences, Oujda, Morocco
ibtissamberriah@gmail.com

³ Associate Professor, Mohammed First University, Faculty of Letters and Human Sciences, Oujda, Morocco
akaouachi@hotmail.com

**L'intelligence artificielle au service de l'évaluation des apprentissages en sciences:
Apports et défis à travers une revue de littérature (2020-2025)**

Chaymae Bouzidi, Ibtissam Benrriah, Abdelali Kaaouachi

Resumé

Cette étude vise à examiner l'utilisation de l'intelligence artificielle (IA) dans l'évaluation des apprentissages en sciences, en identifiant les outils et techniques actuellement employés, les types d'apprentissages pouvant être efficacement évalués, ainsi que les avantages et les limites associés aux pratiques d'évaluation basées sur l'IA. Une revue de littérature systématique a été conduite, portant sur l'analyse de 20 études récentes publiées entre 2020 et 2025, centrées sur les applications de l'IA dans le contexte de l'évaluation des apprentissages en sciences. Les résultats montrent que l'IA permet d'évaluer efficacement divers types d'apprentissages scientifiques, notamment la compréhension conceptuelle, les compétences pratiques, l'argumentation scientifique et la résolution de problèmes. Par rapport aux méthodes traditionnelles, l'IA offre des avantages significatifs tels qu'un feedback en temps réel, une personnalisation de l'apprentissage et une réduction de la charge de travail des enseignants. Cependant, des défis persistent, notamment les biais algorithmiques, la complexité des réponses des apprenants et les préoccupations éthiques. L'intégration réfléchie de l'IA dans l'évaluation des apprentissages en sciences pourrait transformer significativement l'éducation. Néanmoins, une attention particulière doit être portée afin de garantir une utilisation éthique et équitable dans les contextes éducatifs.

Mots-clés : Intelligence artificielle (IA), Évaluation des apprentissages, Apprentissage des sciences.

Introduction

L'évaluation des apprentissages constitue une composante fondamentale du processus éducatif, particulièrement dans l'enseignement des sciences où la compréhension conceptuelle, la pensée critique et les compétences expérimentales doivent être mesurées avec précision (Black & Wiliam, 2010). Dans un contexte éducatif en constante évolution, marqué par la numérisation croissante des ressources pédagogiques et l'émergence de nouvelles approches d'enseignement, les méthodes traditionnelles d'évaluation se révèlent parfois insuffisantes pour capturer la complexité des apprentissages scientifiques (Pellegrino & Quellmalz, 2010). L'intelligence artificielle (IA) émerge comme une technologie prometteuse susceptible de transformer radicalement les pratiques évaluatives en offrant des possibilités d'évaluation personnalisée, continue et multidimensionnelle (Martínez-Comesaña et al., 2023).

Au cours des dernières années, l'intégration de l'IA dans l'évaluation des apprentissages a connu une progression significative. Des systèmes d'évaluation automatisée utilisant le traitement du langage naturel peuvent désormais analyser les réponses textuelles des apprenants (Zhai et al., 2022). Tandis que les algorithmes d'apprentissage automatique permettent d'identifier des modèles complexes dans les données d'apprentissage et de fournir des rétroactions personnalisées (Hooda et al., 2022). Dans le domaine spécifique des sciences, des applications d'IA ont été développées pour évaluer la compréhension conceptuelle, analyser les démarches expérimentales et suivre le développement des compétences scientifiques (Bewersdorff et al., 2023; Gohari et al., 2023; Nehm, 2024).

Malgré ces avancées, l'application de l'IA à l'évaluation des apprentissages en sciences présente encore de nombreux défis. Les questions d'équité algorithmique, de validité des évaluations automatisées et d'interprétabilité des résultats demeurent préoccupantes (Ouyang et al., 2023). Par ailleurs, l'intégration de ces technologies dans les pratiques pédagogiques existantes nécessite une adaptation considérable des enseignants et des institutions (Owan et al., 2023). La littérature scientifique actuelle révèle également un manque de consensus sur les cadres théoriques et méthodologiques permettant d'évaluer l'efficacité de ces systèmes d'IA dans différents contextes d'apprentissage des sciences (Pellegrino, 2024).

Bien que plusieurs études aient examiné l'utilisation de l'IA dans l'éducation en général ou dans certains aspects spécifiques de l'évaluation, il y a peu des études portant spécifiquement sur l'application de l'IA à l'évaluation des apprentissages en sciences. Dans ce contexte, la présente revue de littérature vise donc à combler cette lacune en proposant une analyse approfondie des recherches existantes sur l'utilisation de l'IA pour l'évaluation des apprentissages en sciences. Cette revue de littérature est guidée par la question de recherche suivante : quelles sont les applications de l'intelligence artificielle dans l'évaluation des apprentissages en sciences selon la littérature scientifique récente ? Cette question principale se décline en plusieurs questions spécifiques :

1. Quels outils et techniques d'IA sont actuellement utilisés pour l'évaluation des apprentissages en sciences ?

2. Quels types d'apprentissages scientifiques ces outils permettent-ils d'évaluer efficacement ?
3. Quels sont les avantages démontrés de ces applications par rapport aux méthodes d'évaluation traditionnelles ?
4. Quels sont les principales limites et défis identifiés dans l'utilisation de l'IA pour l'évaluation en sciences ?

Cet article est structuré en cinq sections. Les deux premières sections sont consacrées au cadre théorique : nous y introduisons l'évaluation des apprentissages en sciences et l'IA dans l'éducation. Ensuite, nous expliquons notre méthode pour réaliser cette revue. Enfin, nous présentons les principaux résultats pour étayer la discussion, les conclusions et les tendances futures.

1. Évaluation des apprentissages en sciences

L'évaluation des apprentissages constitue une dimension fondamentale du processus éducatif en sciences, servant non seulement à mesurer les acquis des apprenants, mais également à orienter l'enseignement et à favoriser le développement des compétences scientifiques (Raulin, 2017). Dans le contexte de l'enseignement des sciences, l'évaluation revêt une importance particulière en raison de la nature multidimensionnelle des apprentissages visés, qui englobent à la fois des connaissances conceptuelles, des compétences procédurales et des attitudes (Liu, 2009).

Traditionnellement, l'évaluation des apprentissages en sciences s'est appuyée sur diverses méthodologies, incluant les tests standardisés, les examens écrits, les rapports de laboratoire et les présentations orales (Wragg, 2001). Ces approches ont permis d'évaluer certains aspects des apprentissages scientifiques, notamment la mémorisation de faits et la compréhension de concepts fondamentaux (Lepareur et al., 2023). Cependant, comme le soulignent Pellegrino et al. (2022), ces méthodes traditionnelles présentent des limites significatives, particulièrement dans leur capacité à capturer la complexité du raisonnement scientifique, la démarche d'investigation et la pensée critique.

Les défis spécifiques à l'évaluation des compétences scientifiques sont multiples. Premièrement, la nature intégrée des compétences scientifiques rend difficile leur évaluation isolée (Osborne, 2015). Par exemple, évaluer la capacité d'un élève à concevoir une expérience implique simultanément d'évaluer sa compréhension conceptuelle, ses compétences en matière de contrôle de variables et sa capacité à formuler des hypothèses testables. Deuxièmement, l'évaluation authentique des pratiques scientifiques nécessite des tâches complexes et chronophages, difficiles à mettre en œuvre dans les contraintes temporelles et logistiques des environnements scolaires (Ruiz-Primo & Furtak, 2007). Troisièmement, la subjectivité inhérente à l'évaluation de certaines compétences scientifiques, comme l'argumentation ou la modélisation, pose des défis de fiabilité et de validité (Gotwals & Birmingham, 2016).

Par ailleurs, les évolutions récentes dans la conceptualisation de l'apprentissage des sciences, notamment l'accent mis sur les pratiques scientifiques et l'alphabétisation scientifique, ont accentué le décalage entre les objectifs d'apprentissage contemporains et les méthodes

d'évaluation traditionnelles (Krajcik & Delen, 2017). Comme l'expliquent Harris et al. (, les évaluations traditionnelles tendent à privilégier les connaissances déclaratives au détriment des compétences procédurales et épistémiques, pourtant essentielles à la formation scientifique. Face à ces défis, la communauté éducative a progressivement reconnu la nécessité de développer des approches d'évaluation plus sophistiquées, capables de capturer la richesse et la complexité des apprentissages scientifiques (Zhai, 2023).

2. L'intelligence artificielle dans le domaine de l'éducation

L'intégration de l'intelligence artificielle dans le domaine éducatif représente une évolution significative qui transforme progressivement les pratiques d'enseignement et d'apprentissage (Holmes et al., 2019). L'IA, définie comme la capacité des systèmes informatiques à effectuer des tâches qui nécessiteraient normalement l'intelligence humaine, englobe diverses technologies telles que l'apprentissage automatique, le traitement du langage naturel, la vision par ordinateur et les systèmes experts (De Spiegeleire et al., 2017).

Dans le contexte éducatif, les applications de l'IA se sont multipliées au cours de la dernière décennie, couvrant un large spectre d'usages pédagogiques. Les systèmes tutoriels intelligents (STI) constituent l'une des applications les plus développées, offrant un accompagnement personnalisé aux apprenants en adaptant dynamiquement le contenu et les stratégies pédagogiques en fonction de leurs besoins spécifiques (Mousavinasab et al., 2021). Parallèlement, les chatbots éducatifs et les agents conversationnels, s'appuyant sur les avancées en traitement du langage naturel, facilitent les interactions pédagogiques en fournissant un soutien immédiat et personnalisé aux apprenants (Liu et al., 2022).

L'analyse de l'apprentissage (learning analytics) constitue également un domaine d'application prometteur, où l'IA permet d'analyser de vastes ensembles de données éducatives pour identifier des modèles d'apprentissage, prédire les performances futures et informer les interventions pédagogiques (Chen et al., 2020). Ces approches analytiques offrent aux enseignants et aux institutions des remarques pour optimiser les parcours d'apprentissage et réduire les risques d'échec scolaire (Heeg & Avraamidou, 2023).

Les avantages de l'IA en éducation sont multiples. Premièrement, elle permet une personnalisation sans précédent de l'apprentissage, adaptant le contenu, le rythme et les approches pédagogiques aux besoins individuels des apprenants (Martin et al., 2024). Deuxièmement, elle offre des possibilités d'automatisation de tâches répétitives, libérant ainsi du temps pour les enseignants qui peuvent se concentrer sur des interactions pédagogiques à plus haute valeur ajoutée (Zhai et al., 2023). Enfin, elle permet une évaluation continue et formative, fournissant des retours immédiats et personnalisés aux apprenants (Swiecki et al., 2022).

Malgré ces avantages considérables, l'intégration de l'IA dans l'éducation soulève également des préoccupations et présente des limites. Les questions d'équité algorithmique et de biais constituent une préoccupation majeure, car les systèmes d'IA peuvent perpétuer ou amplifier les inégalités existantes si leurs algorithmes sont entraînés sur des données biaisées (Heeg &

Avraamidou, 2023). La protection des données et le respect de la vie privée représentent également des enjeux cruciaux, particulièrement lorsque les systèmes collectent et analysent des données sensibles sur les apprenants (Bond et al., 2024). Par ailleurs, la dépendance excessive à la technologie risque de diminuer l'autonomie des apprenants et de réduire les interactions humaines essentielles au développement socio-émotionnel (Charles et al., 2025).

Le rôle croissant de l'IA dans l'enseignement et l'apprentissage s'inscrit dans une évolution plus large des paradigmes éducatifs, caractérisée par une transition vers des approches centrées sur l'apprenant, l'apprentissage personnalisé et l'évaluation continue (Martin et al., 2024). Comme le soulignent Liu et Zhong (2024), cette évolution nécessite non seulement des innovations technologiques, mais également des transformations dans les pratiques pédagogiques, les politiques éducatives et la formation des enseignants.

3. L'IA pour l'évaluation des apprentissages

L'application de l'intelligence artificielle à l'évaluation des apprentissages représente un domaine où les avancées technologiques ouvrent de nouvelles perspectives pour surmonter les limitations des approches évaluatives traditionnelles (Martínez-Comesaña et al., 2023). Dans ce contexte, l'IA transforme progressivement les pratiques d'évaluation en permettant des approches plus personnalisées, continues et multidimensionnelles.

Les systèmes d'évaluation basés sur l'IA s'appuient sur diverses technologies et approches. Le traitement automatique du langage naturel (TALN) permet d'analyser et d'évaluer les réponses textuelles des apprenants, facilitant ainsi l'évaluation de compétences complexes comme l'argumentation scientifique ou l'explication de phénomènes (Zhai, 2023).

Les techniques d'apprentissage automatique, notamment les réseaux de neurones profonds, sont également exploitées pour analyser des données multimodales issues des interactions des apprenants avec les environnements d'apprentissage numériques (Latif, Lee, et al., 2024). Les systèmes d'évaluation adaptative, quant à eux, ajustent dynamiquement la difficulté des questions en fonction des performances antérieures de l'apprenant, permettant ainsi une évaluation plus précise et efficiente des niveaux de compétence (Sein Minn, 2022).

En termes de performances, plusieurs études comparatives ont démontré que les systèmes d'évaluation basés sur l'IA peuvent atteindre des niveaux de fiabilité comparables à ceux des évaluateurs humains dans certains contextes. Par exemple, Rudolph et al. (2023) ont rapporté que leur système automatisé d'évaluation de réponses en sciences atteignait un accord interévaluateurs (entre le système et des experts humains) similaire à celui observé entre différents évaluateurs humains. De même, Gao et al. (2025) ont démontré que les systèmes d'IA peuvent évaluer efficacement les explications scientifiques des élèves avec une précision comparable à celle des enseignants. Ces systèmes présentent également l'avantage de pouvoir traiter rapidement de grandes quantités de données, permettant ainsi une évaluation plus fréquente et à plus grande échelle (Alamäki et al., 2024).

Malgré ces avancées prometteuses, l'utilisation de l'IA pour l'évaluation des apprentissages soulève plusieurs défis méthodologiques et éthiques. Sur le plan méthodologique, la validité des évaluations automatisées reste une préoccupation majeure (Gardner et al., 2021). Par ailleurs, la transparence et l'explicabilité des algorithmes d'évaluation constituent un enjeu crucial pour garantir la confiance des parties prenantes et permettre une interprétation adéquate des résultats (Owan et al., 2023).

Sur le plan éthique, plusieurs questions méritent une attention particulière. Les risques de biais algorithmiques peuvent conduire à des évaluations inéquitables, désavantageant certains groupes d'apprenants en fonction de caractéristiques comme la langue maternelle, le contexte socioculturel ou le style d'apprentissage (Gonçalves et al., 2024). La protection des données personnelles et le consentement éclairé des apprenants représentent également des préoccupations majeures, particulièrement lorsque les systèmes collectent des données comportementales détaillées (Yaneva & von Davier, 2023). Enfin, l'autonomie professionnelle des enseignants pourrait être menacée si les systèmes automatisés remplacent plutôt que de compléter leur expertise évaluative (Gardner et al., 2021).

Méthodologie

La méthodologie adoptée pour cette revue de littérature s'est articulée autour de plusieurs étapes visant à garantir l'exhaustivité et la pertinence de l'analyse.

1. Recherche documentaire

La recherche documentaire a été effectuée dans cinq bases de données scientifiques reconnues pour leur couverture en sciences de l'éducation et en technologies éducatives : Scopus, Springer, Web of Science, ScienceDirect et Research4Life. Les mots-clés principaux utilisés pour la recherche étaient "Assessment" AND "Artificial Intelligence" AND "Education". Pour cibler plus précisément notre sujet, nous avons également employé des termes complémentaires tels que "Science education", "STEM". Cette stratégie a permis d'identifier un corpus initial de 107 articles potentiellement pertinents.

2. Critères de sélection

Après l'identification du corpus initial, un processus de sélection a été mis en place afin d'affiner la pertinence des articles retenus. Les critères d'inclusion exigeaient que les publications soient parues entre 2020 et 2025, rédigées en anglais ou en français, qu'elles mettent explicitement l'accent sur l'utilisation de l'intelligence artificielle pour l'évaluation des apprentissages en sciences, et qu'elles soient accessibles en texte intégral. À l'inverse, ont été exclus de la sélection les articles ne portant pas spécifiquement sur l'évaluation des apprentissages, ceux n'abordant pas l'enseignement des sciences, ainsi que les publications issues de revues jugées de qualité insuffisante. À l'issue de ce processus, 20 articles publiés entre 2020 et 2025 ont été retenus pour une analyse approfondie.

3. Analyse qualitative des articles

Chaque article retenu a fait l'objet d'une lecture intégrale et d'un codage systématique, dans des catégories prédéfinies, basées sur le cadre théorique de l'étude : techniques et outils d'intelligence artificielle mobilisés, niveaux d'enseignement concernés, types d'apprentissages ciblés, avantages, limites soulevées. Le logiciel NVivo 14 a été utilisé pour organiser, catégoriser et analyser les données extraites. Pour garantir la fiabilité de l'analyse, un échantillon de cinq articles a été codé indépendamment par un autre chercheur. Les divergences d'interprétation ont été discutées et résolues, ce qui a permis d'assurer une application cohérente des codes à l'ensemble du corpus.

5. Synthèse et organisation des résultats

Les résultats ont été synthétisés et organisés selon les thématiques identifiées lors de l'analyse, offrant ainsi une vue d'ensemble des applications actuelles, des opportunités et des défis liés à l'utilisation de l'intelligence artificielle pour l'évaluation des apprentissages en sciences.

Résultats et Discussion

Cette section présente les résultats de l'analyse des 20 articles examinés, en mettant en lumière les outils et techniques d'intelligence artificielle (IA) utilisés pour l'évaluation des apprentissages en sciences, les types d'apprentissages évalués, les avantages par rapport aux méthodes traditionnelles, ainsi que les limites et défis associés à l'utilisation de l'IA dans ce domaine.

Le tableau 1 suivant présente les principales données des études sélectionnées : les auteurs, l'année de publication, la revue, le pays d'origine des auteurs, leur affiliation professionnelle, ainsi que la matière ou le niveau d'enseignement dans lesquels l'IA a été utilisée sont mis en évidence.



(Gao et al., 2025)	2025	Journal of Research in Science Teaching	États-unites	Mathématiques, sciences et sciences sociales	Enseignement primaire et secondaire Mathématiques et sciences
(Pellegrino, 2024)	2024	Oxford University Press	États-unites	Technologie éducative	Enseignement supérieur STEM
(Ouyang et al., 2023)	2023	Journal for STEM Education Research	États-unites	Ingénierie, technologie éducative	STEM
(Zhai & Nehm, 2023)	2023	Journal of Research in Science Teaching	États-unites	Mathématiques, sciences et sciences sociales	Enseignement primaire et secondaire Mathématiques et sciences
(Zhai, 2023)	2023	Oxford University Press	États-unites	Mathématiques, sciences et sciences sociales	Enseignement primaire et secondaire Sciences
(Nehm, 2024)	2024	Oxford University Press	États-unites	Écologie et évolution, enseignement des sciences	Biologie
(Latif, Zhai, et al., 2024)	2024	Oxford University Press	États-unites	Mathématiques, sciences et sciences sociales	Enseignement primaire et secondaire STEM
(Vittorini et al., 2021)	2021	International Journal of Artificial Intelligence in Education	Italia	Médecine et chirurgie	Enseignement supérieur Médecine et chirurgie
(Lee et al., 2024)	2024	Computers and Education: Artificial Intelligence	États-unites	Mathématiques, sciences et sciences sociales	Enseignement secondaire collégial Sciences
(Zhai et al., 2022)	2022	Journal of Research in Science Teaching	États-unites	Mathématiques, sciences et sciences sociales	Enseignement secondaire collégial Sciences
(Bewersdorff et al., 2023)	2023	Computers and Education: Artificial Intelligence	Allemagne	Sciences sociales et technologie	Enseignement secondaire Mathématiques et sciences
(Chan et al., 2025)	2025	Computers and Education: Artificial Intelligence	Singapore	Intelligence artificielle, technologie éducative, chimie	Enseignement secondaire qualifiant Chimie, physique et mathématique

Auteurs	Année de publication	Revue	Pays	Affiliation professionnelle	Matière ou le niveau d'enseignement
(Haudek & Zhai, 2024)	2024	International Journal of Artificial Intelligence in Education	États-unites	Biochimie, biologie moléculaire, STEM	Enseignement secondaire collégial Sciences
(Latif & Zhai, 2024)	2024	Computers and Education: Artificial Intelligence	États-unites	Mathématiques, sciences et sciences sociales	Enseignement secondaire Sciences
(Latif, Lee, et al., 2024)	2024	Computer Science, Éducation	Allemagne	Mathématiques, sciences et sciences sociales	Enseignement secondaire collégial Sciences
(Nyaaba et al., 2024)	2024	Éducation Sciences	États-unites	STEM, Science d'éducation	Enseignement primaire et secondaire Sciences STEM
(Zhai et al., 2021)	2021	Journal of Science, Education and Technology	États-unites	Mathématiques, sciences et sciences sociales	Enseignement primaire et secondaire Sciences
(Zhai, 2021)	2021	Journal of Science, Education and Technology	États-unites	Mathématiques, sciences et sciences sociales	Enseignement primaire et secondaire Sciences
(Bedi et al., 2025)	2025	Pacific Symposium on Biocomputing	États-unites	Science des données biomédicales	Enseignement supérieur Médecine et chirurgie

Tableau 1. Les principales données des études sélectionnées

1. Outils et Techniques d'IA Utilisés pour l'Évaluation des Apprentissages en Sciences

Les recherches analysées révèlent plusieurs catégories d'outils et de techniques d'IA qui transforment cette pratique. Ces outils et techniques d'IA actuellement utilisés pour l'évaluation des apprentissages en sciences reposent principalement sur les modèles de langage larges (LLM) pour la correction automatique, la génération d'items et l'analyse d'erreurs, et sur l'apprentissage automatique pour l'évaluation de tâches complexes, l'analyse prédictive et l'adaptation des évaluations et des plateformes intégrées facilitant le feed-back immédiat, la personnalisation et l'équité.

Automatisation de la correction et du feed-back

Les modèles de langage larges (LLM) tels que GPT-3.5, GPT-4 et BERT, ainsi que leurs variantes spécialisées comme G-SciEdBER, sont désormais largement utilisés pour la correction automatique des réponses écrites en sciences (Latif, Lee, et al., 2024; Latif & Zhai, 2024). Les

stratégies de prompt engineering, notamment le chain-of-thought (CoT), améliorent la précision et l'explicabilité du scoring, surtout lorsqu'elles sont combinées à des rubriques d'évaluation (Lee et al., 2024). Ces modèles se basent sur des corpus spécifiques d'élèves et permettent d'atteindre des niveaux de concordance avec les évaluateurs humains jugés satisfaisants (Haudek & Zhai, 2024). De plus, des plateformes telles qu'AI-Scorer ou EvoGrader offrent des environnements d'évaluation formative automatisée, fournissant un feed-back immédiat et personnalisé aux élèves et aux enseignants, facilitant ainsi l'ajustement pédagogique en temps réel (Latif, Lee, et al., 2024; Nehm, 2024).

Des systèmes d'apprentissage automatique ML sont également employés pour l'évaluation de productions complexes, telles que les modèles scientifiques dessinés et leurs descriptions, ou les protocoles expérimentaux (Bewersdorff et al., 2023; Zhai et al., 2022). Les algorithmes supervisés (SVM, Random Forest, réseaux de neurones) sont entraînés à partir de notations humaines pour automatiser la détection d'erreurs, l'identification de concepts clés ou l'évaluation de la qualité des argumentations scientifiques (Haudek & Zhai, 2024).

Génération automatisée d'items d'évaluation

Les LLM tels que GPT-3.5, GPT-4 sont utilisés pour générer automatiquement des items d'évaluation en sciences, y compris des questions à choix multiples, des questions ouvertes ou des exercices de programmation (Bedi et al., 2025; Chan et al., 2025). L'utilisation de stratégies de prompt engineering, telles que le chain-of-thought et l'intégration d'exemples, améliore la qualité et la pertinence des questions générées, tout en limitant les risques d'erreurs ou de biais (Chan et al., 2025). Des outils comme le CRSciA-Generator exploitent les capacités des LLM pour produire des items adaptés aux contextes culturels des élèves, favorisant ainsi l'équité et la pertinence des évaluations dans des environnements multiculturels (Nyaaba et al., 2024).

Systèmes adaptatifs et learning analytics

Les systèmes adaptatifs, intégrant des algorithmes de clustering ou de knowledge tracing, ajustent la difficulté et la nature des tâches en fonction des performances de l'élève (Ouyang et al., 2023). L'apprentissage supervisé (forêts aléatoires, XGBoost, réseaux de neurones profonds) est utilisé pour prédire les performances futures des élèves à partir de leurs traces d'apprentissage, permettant d'identifier les élèves à risque et d'orienter les interventions pédagogiques (Pellegrino, 2024).

2. Types d'Apprentissages scientifiques évalués

Les résultats montrent que les outils d'IA et de ML permettent d'évaluer efficacement les connaissances disciplinaires et leur mobilisation en contexte, les pratiques scientifiques telles que la modélisation, l'expérimentation, l'argumentation, la résolution de problèmes et les compétences transversales dans des contextes variés.

Modélisation scientifique : Les systèmes de ML peuvent scorer automatiquement des modèles dessinés et leurs descriptions écrites, permettant d'évaluer la capacité des élèves à représenter, expliquer et prédire des phénomènes (Zhai et al., 2022).

Expérimentation et raisonnement expérimental : Les LLMs détectent les erreurs dans les protocoles expérimentaux, identifient les failles de raisonnement et fournissent un retour personnalisé sur la conception et l'interprétation d'expériences (Bewersdorff et al., 2023).

Argumentation et explication : Les modèles automatisés évaluent la structure, la complexité et la diversité des arguments scientifiques, tout en identifiant les niveaux de progression dans l'argumentation (Haudek & Zhai, 2024).

Résolution de problèmes et application des connaissances : Les outils évaluent la capacité à résoudre des problèmes complexes, notamment en mathématiques, physique, chimie et biologie, à travers des réponses ouvertes ou des exercices de codage (Vittorini et al., 2021).

Compétences transversales : L'IA permet d'évaluer la pensée critique, la créativité, la capacité à relier différents concepts et à transférer des connaissances entre disciplines (Ouyang et al., 2023; Zhai, 2021).

3. Avantages des Applications d'IA par rapport aux Méthodes traditionnelles

L'analyse des vingt articles sélectionnés met en évidence une série d'avantages démontrés des applications d'intelligence artificielle (IA) dans l'évaluation des apprentissages en sciences, en comparaison avec les méthodes traditionnelles.

Efficacité et gain de temps

Un consensus fort émerge sur la capacité des systèmes d'IA à automatiser la correction et la génération d'items d'évaluation, réduisant significativement le temps et la charge de travail des enseignants (Lee et al., 2024; Vittorini et al., 2021; Zhai et al., 2022). Par exemple, l'utilisation de systèmes de correction automatique (tels qu'AI-Scorer ou EvoGrader) permet d'obtenir des résultats en temps réel, facilitant ainsi la rétroaction immédiate aux élèves et la prise de décision pédagogique rapide (Lee et al., 2024; Nehm, 2024). De plus, la génération automatique de questions par LLM accélère la création de banques d'items, tout en maintenant une qualité comparable à celle des items humains (Bedi et al., 2025; Chan et al., 2025).

Amélioration de la qualité et de l'objectivité de l'évaluation

Les études montrent que les systèmes d'IA atteignent des niveaux de concordance élevés avec les évaluations humaines, tout en offrant une plus grande stabilité et cohérence intertâches (Lee et al., 2024; Zhai et al., 2020). Les algorithmes sont moins sensibles aux variations de contexte ou à la subjectivité des correcteurs, ce qui réduit la variance non pertinente (construct-irrelevant variance) et améliore la validité de l'évaluation (Zhai et al., 2021). L'utilisation de modèles spécialisés, comme G-SciEdBERT pour l'allemand, montre également une amélioration significative de la précision par rapport aux modèles généralistes (Latif, Lee, et al., 2024).

Personnalisation et feed-back immédiat

L'IA permet de fournir un feed-back personnalisé, détaillé et immédiat aux élèves, favorisant l'apprentissage autorégulé et l'engagement (Latif, Zhai, et al., 2024; Zhai & Nehm, 2023). Les systèmes peuvent identifier des erreurs spécifiques dans les productions des élèves, générer des conseils adaptés et aider les enseignants à cibler les besoins individuels ou collectifs de la classe (Bewersdorff et al., 2023; Ouyang et al., 2023).

Diversification des formats et des modalités d'évaluation

Les applications d'IA facilitent l'évaluation de compétences complexes et multidimensionnelles, difficiles à appréhender par des méthodes traditionnelles (Zhai et al., 2023). Elles permettent l'analyse automatique de réponses ouvertes, de modèles scientifiques dessinés, d'argumentations, ou encore de protocoles expérimentaux (Haudek & Zhai, 2024). Les tâches multimodales et interactives proposées favorisent l'alignement avec les objectifs des curriculums de nouvelle génération (NGSS), en évaluant non seulement les connaissances, mais aussi les pratiques et les compétences transversales (Gao et al., 2025).

Accessibilité, équité et adaptation culturelle

Plusieurs travaux soulignent que l'IA peut contribuer à une évaluation plus équitable, en réduisant certains biais humains et en s'adaptant à la diversité linguistique et culturelle (Latif, Lee, et al., 2024). Des outils comme le CRSciA-Generator permettent de générer des items contextualisés et culturellement adaptés, favorisant l'inclusion d'élèves issus de milieux variés (Nyaaba et al., 2024). De plus, l'automatisation facilite l'accès à des évaluations de qualité dans des contextes à grande échelle (MOOCs, évaluations nationales ou internationales) (Zhai & Nehm, 2023).

Innovation pédagogique et transformation des pratiques

L'intégration de l'IA dans l'évaluation favorise l'innovation pédagogique, en permettant de concevoir des tâches d'évaluation inédites, interactives et authentiques (Gao et al., 2025). Les enseignants peuvent s'appuyer sur les analyses de données produites par l'IA pour ajuster leur enseignement, différencier les parcours et s'engager dans des pratiques d'évaluation formative plus riches (Latif, Lee, et al., 2024). L'IA ouvre également la voie à l'évaluation de compétences transversales et à la prise en compte de la progression des apprentissages sur le long terme (Pellegrino, 2024).

4. Limites et Défis de l'Utilisation de l'IA pour l'Évaluation en Sciences

L'analyse des 20 articles révèle que, malgré le potentiel transformateur de l'intelligence artificielle (IA) dans l'évaluation en sciences, plusieurs limites et défis majeurs persistent.

Validité des évaluations automatisées

Plusieurs études soulignent que l'IA, notamment par l'apprentissage automatique, introduit de nouveaux risques de validité. Parmi ceux-ci figurent la possibilité de mal représenter le construit

d'intérêt, le non-alignement entre les scores générés par la machine et les objectifs d'apprentissage, ou encore la difficulté à extrapoler les résultats à d'autres contextes ou populations (Pellegrino, 2024; Zhai et al., 2022). Les modèles peuvent être sensibles à des variables non pertinentes, menaçant la validité des inférences tirées des scores.

Limites face à la complexité des tâches scientifiques

Les performances des modèles d'IA diminuent lorsque les tâches d'évaluation sont complexes, multidimensionnelles ou peu structurées (Haudek & Zhai, 2024). Par exemple, l'évaluation de la modélisation scientifique ou de l'argumentation écrite reste difficile à automatiser avec une précision équivalente à celle des experts humains, surtout lorsque les réponses présentent une grande diversité ou des structures argumentatives complexes (Gao et al., 2025).

Biais algorithmiques

Les systèmes d'IA peuvent reproduire ou amplifier des biais présents dans les données d'entraînement, affectant différemment les élèves selon leur origine linguistique, culturelle ou socio-économique (Nyaaba et al., 2024). Les modèles généralistes, entraînés sur des corpus non contextualisés, peinent à prendre en compte la diversité des styles d'écriture ou des référents culturels (Latif, Lee, et al., 2024). Dans la génération automatique d'items d'évaluation, les modèles de langage peuvent produire des questions erronées, incohérentes ou non alignées avec les objectifs pédagogiques (Bedi et al., 2025). La vérification humaine reste souvent nécessaire pour garantir la qualité et la validité des items générés.

Problèmes d'explicabilité et de transparence

L'explicabilité des décisions prises par les modèles d'IA reste un défi majeur (Lee et al., 2024). Les enseignants et les élèves peuvent avoir du mal à comprendre les critères de notation ou les raisons d'un score attribué, ce qui limite la confiance et l'acceptation des systèmes automatisés (Chan et al., 2025).

Questions éthiques et de confidentialité

L'utilisation de l'IA soulève des préoccupations éthiques concernant la confidentialité des données des étudiants et l'utilisation de ces données à des fins d'évaluation. Nyaaba et al. (2024) mettent en garde contre le risque de collecte excessive de données et soulignent l'importance de garantir la protection des informations personnelles des étudiants.

Discussion

Les outils d'IA, tels que les systèmes de génération automatique des items et les modèles de correction automatique, représentent une avancée significative dans le domaine de l'évaluation. Ces technologies permettent non seulement de gagner du temps dans la création et la correction des évaluations, mais elles offrent également une personnalisation qui était difficile à atteindre avec les méthodes traditionnelles. Par exemple, le système QUEST-AI illustre comment l'IA peut générer des questions adaptées au niveau de compétence des étudiants, ce qui est essentiel pour

répondre aux besoins d'une population étudiante diversifiée (Bedi et al., 2025). L'évaluation des erreurs des élèves, comme le montre l'étude de Bewersdorff et al. (2023), souligne également l'importance de fournir un feed-back constructif. En identifiant les erreurs spécifiques, l'IA permet aux élèves de comprendre leurs lacunes et d'améliorer leur apprentissage. Cela est particulièrement pertinent dans les disciplines scientifiques où la compréhension des concepts fondamentaux est importante pour le succès futur.

Notamment, l'IA est capable d'évaluer une variété d'apprentissages scientifiques, allant de la compréhension conceptuelle à la résolution de problèmes. Cette capacité à évaluer des compétences complexes est essentielle dans un monde où les étudiants doivent non seulement mémoriser des faits, mais aussi appliquer leurs connaissances dans des contextes variés. Par exemple, l'évaluation du raisonnement scientifique montre que l'IA peut analyser la logique et la structure des réponses des étudiants, ce qui est fondamental pour développer des compétences critiques et analytiques (Haudek & Zhai, 2024). Cependant, il est important de noter que l'évaluation des compétences pratiques et des réponses ouvertes reste un défi. Bien que l'IA puisse fournir des évaluations objectives, elle peut parfois manquer de nuance dans l'analyse des réponses créatives ou des approches innovantes. Cela soulève des questions sur la manière dont l'IA peut être intégrée dans des évaluations qui nécessitent une appréciation humaine.

Les avantages des applications d'IA par rapport aux méthodes traditionnelles sont présentés principalement par le feed-back en temps réel, la personnalisation de l'apprentissage et la réduction de la charge de travail des enseignants, des éléments qui peuvent transformer l'expérience d'apprentissage. Les systèmes d'évaluation assistés par IA permettent aux enseignants de se concentrer sur l'interaction avec les étudiants, plutôt que sur la correction des travaux, ce qui peut améliorer la qualité de l'enseignement (Gao et al., 2025). Cependant, il est essentiel de ne pas perdre de vue que l'IA ne doit pas remplacer l'interaction humaine dans l'évaluation. Les enseignants jouent un rôle essentiel dans l'interprétation des résultats et dans la fourniture d'un soutien émotionnel et pédagogique aux étudiants (Latif, Lee, et al., 2024). L'IA doit être considérée comme un outil complémentaire qui enrichit l'expérience d'apprentissage, plutôt qu'un substitut à l'enseignement traditionnel.

Les limites et défis identifiés dans l'utilisation de l'IA pour l'évaluation en sciences soulèvent des préoccupations importantes. Le biais des algorithmes, la complexité des réponses et les questions éthiques liées à la confidentialité des données sont des enjeux qui nécessitent une attention particulière (Nyaaba et al., 2024). Ouyang et al. (2023) mettent en évidence que les modèles d'IA peuvent reproduire des biais présents dans les données d'entraînement, ce qui peut affecter l'équité des évaluations. Il est donc essentiel de s'assurer que les données utilisées pour entraîner ces modèles sont représentatives de la diversité des étudiants.

De plus, l'acceptation par les enseignants et les étudiants est un facteur clé pour le succès de l'intégration de l'IA dans l'évaluation. La formation des enseignants et la sensibilisation des étudiants aux avantages et aux limites de l'IA sont importantes pour garantir une adoption réussie (Zhai & Nehm, 2023). Pellegrino (2024) souligne que la résistance au changement peut freiner

l'innovation dans l'éducation, et il est donc impératif de créer un environnement où l'IA est perçue comme un atout.

Conclusion

L'intelligence artificielle (IA) représente une avancée significative dans le domaine de l'évaluation des apprentissages en sciences, offrant des outils et des techniques qui transforment la manière dont les compétences des élèves sont mesurées et développées. Cette revue de littérature a mis en lumière une variété d'applications de l'IA, allant de la génération automatique des items à l'évaluation des erreurs des élèves, en passant par des systèmes d'évaluation multimodaux et des modèles de correction automatique. Ces innovations permettent non seulement d'améliorer la précision et la rapidité des évaluations, mais aussi de personnaliser l'apprentissage en fonction des besoins individuels des élèves. Les résultats montrent que l'IA peut efficacement évaluer divers types d'apprentissages scientifiques, y compris la compréhension conceptuelle, les compétences pratiques, l'argumentation scientifique et la résolution de problèmes. En fournissant un feed-back en temps réel et en réduisant la charge de travail des enseignants, l'IA contribue à créer un environnement d'apprentissage plus dynamique et interactif. Cependant, l'intégration de l'IA dans l'évaluation des apprentissages en sciences n'est pas sans défis. Les préoccupations concernant les biais algorithmiques, la complexité des réponses, l'acceptation par les enseignants et les élèves, ainsi que les questions éthiques liées à la confidentialité des données doivent être abordées de manière proactive. Il est essentiel de garantir que l'IA soit utilisée de manière éthique et équitable, en complément des méthodes d'évaluation traditionnelles. À l'avenir, l'intégration de l'IA dans l'évaluation des apprentissages en sciences pourrait ouvrir de nouvelles avenues pour l'enseignement et l'apprentissage. Les recherches futures devraient se concentrer sur le développement de modèles d'IA capables de mieux comprendre le contexte et la créativité des réponses des étudiants. De plus, il sera essentiel d'explorer des approches hybrides qui combinent l'IA avec l'évaluation humaine pour garantir une évaluation complète.

Références

- Alamäki, A., Khan, U. A., Kauttonen, J., & Schlögl, S. (2024). An Experiment of AI-Based Assessment : Perspectives of Learning Preferences, Benefits, Intention, Technology Affinity, and Trust. *Education Sciences*, 14(12). Scopus. <https://doi.org/10.3390/educsci14121386>
- Bedi, S., Fleming, S. L., Chiang, C.-C., Morse, K., Kumar, A., Patel, B., Jindal, J. A., Davenport, C., Yamaguchi, C., & Shah, N. H. (2025). QUEST-AI : A System for Question Generation, Verification, and Refinement using AI for USMLE-Style Exams. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 30, 54-69. Scopus.
- Bewersdorff, A., Seßler, K., Baur, A., Kasneci, E., & Nerdel, C. (2023). Assessing student errors in experimentation using artificial intelligence and large language models : A comparative study with human raters. *Computers and Education: Artificial Intelligence*, 5, 100177. <https://doi.org/10.1016/j.caeai.2023.100177>
- Black, P., & William, D. (2010). Inside the Black Box: Raising Standards Through Classroom Assessment. [http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/\[in=epidoc1.in\]/?t2000=022921/\(100\), 80](http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=022921/(100), 80). <https://doi.org/10.1177/003172171009200119>
- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta-systematic review of artificial intelligence in higher education : A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21(1). Scopus. <https://doi.org/10.1186/s41239-023-00436-z>
- Chan, K. W., Ali, F., Park, J., Sham, K. S. B., Tan, E. Y. T., Chong, F. W. C., Qian, K., & Sze, G. K. (2025). Automatic item generation in various STEM subjects using large language model prompting. *Computers and Education: Artificial Intelligence*, 8. Scopus. <https://doi.org/10.1016/j.caeai.2024.100344>
- Charles, K. A., Yousuf, A., Chua, H. C., Matthews, S., Harnett, J., & Hinton, T. (2025). AI in action : Changes to student perceptions when using generative artificial intelligence for the creation of a multimedia project-based assessment. *European Journal of Pharmacology*, 998, 177508. <https://doi.org/10.1016/j.ejphar.2025.177508>
- 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, 100002. <https://doi.org/10.1016/j.caeai.2020.100002>
- De Spiegeleire, S., Maas, M., & Sweijs, T. (2017). *What Is Artificial Intelligence?* (ARTIFICIAL INTELLIGENCE AND THE FUTURE OF DEFENSE, p. 25-42). Hague Centre for Strategic Studies. <https://www.jstor.org/stable/resrep12564.7>
- Gao, Y., Zhai, X., Li, M., Lee, G., & Liu, X. (2025). A Multimodal Interactive Framework for Science Assessment in the Era of Generative Artificial Intelligence. *Journal of Research in Science Teaching*. <https://doi.org/10.1002/tea.70009>
- Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment : « Breakthrough? Or bunkum and ballyhoo? » *JOURNAL OF COMPUTER ASSISTED LEARNING*, 37(5), 1207-1216. <https://doi.org/10.1111/jcal.12577>
- Gohari, V., Sadat, M. K. N., & Ramezaivishki, F. (2023). Competency assessment of biology teaching students of Farhangian University by using an artificial intelligence feedforward neural network. *Fanāvarī-i Āmūzish*, 17(3), 695-708. <https://doi.org/10.22061/tej.2023.9539.2861>
- Goncalves, B. F., Patricio, M. R., & Comiche, A. (2024). The Challenges of Learning Assessment in the Age of Artificial Intelligence. In A. Rocha, H. Adeli, G. Dzemyda, F. Moreira, & A. Poniszewska-Maranda

- (Éds.), *GOOD PRACTICES AND NEW PERSPECTIVES IN INFORMATION SYSTEMS AND TECHNOLOGIES, VOL 4, WORLDCIST 2024* (Vol. 988, p. 23-32). Springer International Publishing Ag.
https://doi.org/10.1007/978-3-031-60224-5_3
- Gotwals, A. W., & Birmingham, D. (2016). Eliciting, Identifying, Interpreting, and Responding to Students' Ideas : Teacher Candidates' Growth in Formative Assessment Practices. *Research in Science Education*, 46(3), 365-388. <https://doi.org/10.1007/s11165-015-9461-2>
 - Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing Knowledge-In-Use Assessments to Promote Deeper Learning. *Educational Measurement: Issues and Practice*, 38(2), 53-67. <https://doi.org/10.1111/emip.12253>
 - Haudek, K. C., & Zhai, X. (2024). Examining the Effect of Assessment Construct Characteristics on Machine Learning Scoring of Scientific Argumentation. *International Journal of Artificial Intelligence in Education*, 34(4), 1482-1509. Scopus. <https://doi.org/10.1007/s40593-023-00385-8>
 - Heeg, D. M., & Avraamidou, L. (2023). The use of Artificial intelligence in school science : A systematic literature review. *Educational Media International*, 60(2), 125-150. <https://doi.org/10.1080/09523987.2023.2264990>
 - Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education. Promise and Implications for Teaching and Learning*.
 - 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(1), 5215722. <https://doi.org/10.1155/2022/5215722>
 - Krajcik, J., & Delen, İ. (2017). Engaging learners in STEM education. *Eesti Haridusteaduste Ajakiri*, 5(1). <https://doi.org/10.12697/eha.2017.5.1.02b>
 - Latif, E., Lee, G., Neuman, K., Kastorff, T., & Zhai, X. (2024). *G-SciEdBERT : A Contextualized LLM for Science Assessment Tasks in German*.
 - Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. *Computers and Education: Artificial Intelligence*, 6, 100210. <https://doi.org/10.1016/j.caeai.2024.100210>
 - Latif, E., Zhai, X., Amerman, H., & He, X. (2024). AI-scorer : An artificial intelligence-augmented scoring and instruction system. In *Uses of Artificial Intelligence in STEM Education* (p. 269-294). Scopus. <https://doi.org/10.1093/oso/9780198882077.003.0013>
 - Lee, G.-G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence*, 6, 100213. <https://doi.org/10.1016/j.caeai.2024.100213>
 - Lepareur, C., Marlot, C., & Monnier, M. D. (2023). Analysis of assessment practices on science education in primary school. *Recherches en Didactiques des Sciences et des Technologies*, (27), 23-50. Scopus. <https://doi.org/10.4000/rdst.4594>
 - Liu, L., Subbareddy, R., & Raghavendra, C. G. (2022). AI Intelligence Chatbot to Improve Students' Learning in the Higher Education Platform. *J. Interconnect. Networks*, 22, 21430321-214303217. <https://doi.org/10.1142/s0219265921430325>
 - Liu, X. (2009). *Essentials of Science Classroom Assessment* (1^{re} éd.). SAGE Publications, Incorporated. <http://gen.lib.rus.ec/book/index.php?md5=6D892EA7A9663E7C616481E80FC4839C>
 - Liu, X., & Zhong, B. (2024). A systematic review on how educators teach AI in K-12 education. *Educational Research Review*, 45. Scopus. <https://doi.org/10.1016/j.edurev.2024.100642>

- Martin, F., Zhuang, M., & Schaefer, D. (2024). Systematic review of research on artificial intelligence in K-12 education (2017–2022). *Computers and Education: Artificial Intelligence*, 6, 100195. <https://doi.org/10.1016/j.caeai.2023.100195>
- Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocarranza-Prado, I., & Kreibel, D. (2023). Impact of artificial intelligence on assessment methods in primary and secondary education : Systematic literature review. *Revista de Psicodidáctica (English ed.)*, 28(2), 93-103. <https://doi.org/10.1016/j.psicoe.2023.06.002>
- Mousavinasab, E., Zarifsanaiy, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems : A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, 29(1), 142-163. <https://doi.org/10.1080/10494820.2018.1558257>
- Nehm, R. H. (2024). AI in biology education assessment : How automation can drive educational transformation. In *Uses of Artificial Intelligence in STEM Education* (p. 38-58). Scopus. <https://doi.org/10.1093/oso/9780198882077.003.0003>
- Nyaaba, M., Zhai, X., & Faison, M. (2024). Generative AI for Culturally Responsive Science Assessment : A Conceptual Framework. *Education Sciences*, 14, 1325. <https://doi.org/10.3390/educsci14121325>
- Osborne, J. (2015). Practical work in science : Misunderstood and badly used? *School Science Review*, 96(357), 16-24.
- Ouyang, F., Dinh, T. A., & Xu, W. (2023). A Systematic Review of AI-Driven Educational Assessment in STEM Education. *Journal for STEM Education Research*, 6(3), 408-426. Scopus. <https://doi.org/10.1007/s41979-023-00112-x>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Pellegrino, J. W. (2024). A New Era for STEM Assessment : Considerations of Assessment, Technology, and Artificial Intelligence. In X. Zhai & J. Krajcik (Éds.), *Uses of Artificial Intelligence in STEM Education* (p. 0). Oxford University Press. <https://doi.org/10.1093/oso/9780198882077.003.0002>
- Pellegrino, J. W., & Quellmalz, E. S. (2010). Perspectives on the Integration of Technology and Assessment. *Journal of Research on Technology in Education*, 43(2), 119-134. <https://doi.org/10.1080/15391523.2010.10782565>
- Raulin, D. (2017). Objets et pratiques de l'évaluation scolaire. *Revue française de pédagogie*, (200), 69-80.
- 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), 342-363. Scopus. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring Teachers' Informal Formative Assessment Practices and Students' Understanding in the Context of Scientific Inquiry. *Journal of Research in Science Teaching*, 44(1), 57-84. <https://doi.org/10.1002/tea.20163>
- Sein Minn. (2022). AI-assisted knowledge assessment techniques for adaptive learning environments. *Computers and Education: Artificial Intelligence*, 3, 100050. <https://doi.org/10.1016/j.caeai.2022.100050>

- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. <https://doi.org/10.1016/j.caeai.2022.100075>
- Vittorini, P., Menini, S., & Tonelli, S. (2021). An AI-Based System for Formative and Summative Assessment in Data Science Courses. *International Journal of Artificial Intelligence in Education*, 31(2), 159-185. <https://doi.org/10.1007/s40593-020-00230-2>
- Wragg, P. E. C. (2001). *Assessment and Learning in the Secondary School* (2^e éd.). RoutledgeFalmer. <http://gen.lib.rus.ec/book/index.php?md5=ff905277622bf2333f04d698e0f7e8d3>
- Yaneva, V., & von Davier, M. (Éds.). (2023). *Advancing Natural Language Processing in Educational Assessment*. Taylor & Francis. <https://doi.org/10.4324/9781003278658>
- 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. <https://doi.org/10.1007/s10956-021-09901-8>
- Zhai, X. (2023). *AI and Machine Learning for Next Generation Science Assessment*.
- Zhai, X., Haudek, K. C., Stuhlsatz, M. A. M., & Wilson, C. (2020). Evaluation of construct-irrelevant variance yielded by machine and human scoring of a science teacher PCK constructed response assessment. *Studies in Educational Evaluation*, 67, 100916. <https://doi.org/10.1016/j.stueduc.2020.100916>
- Zhai, X., He, P., & Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*, 59, 1765-1794. <https://doi.org/10.1002/tea.21773>
- Zhai, X., Krajcik, J., & Pellegrino, J. (2021). On the Validity of Machine Learning-based Next Generation Science Assessments : A Validity Inferential Network. *Journal of Science Education and Technology*, 30. <https://doi.org/10.1007/s10956-020-09879-9>
- Zhai, X., & Nehm, R. H. (2023). AI and formative assessment : The train has left the station. *Journal of Research in Science Teaching*, 60(6), 1390-1398. <https://doi.org/10.1002/tea.21885>
- Zhai, X., Neumann, Knut, & Krajcik, J. (2023). AI for tackling STEM education challenges. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.1183030>