

Les outils/instruments numériques pour l'évaluation des apprentissages

Analyse d'une recherche documentaire en incluant les robots conversationnels

Aristide Tsayem Tchoupou

Doctorante en mesure et évaluation – Université Laval (Canada)

Thérèse Laferrière

FSÉ Département d'études sur l'enseignement et l'apprentissage – Université Laval

Georges-Louis Baron

Laboratoire Education, discours et apprentissages – Université Paris Cité

Numéro spécial 2 / 2023 - T3



RÉSUMÉ L'utilisation des outils/instruments numériques aide à mettre en place de nouvelles pratiques d'enseignement et d'apprentissage. Un défi majeur est d'aligner les nouvelles pratiques d'enseignement et l'évaluation des apprentissages. Pour assurer la validité et la fiabilité des résultats obtenus lors des évaluations, il importe de calibrer correctement les outils de mesure, voire de prévenir le plagiat. De plus, l'émergence des robots conversationnels dans l'éducation pose des questions d'équité d'usage et d'authenticité des productions des élèves. Cet article examine les outils et instruments numériques utilisés pour l'évaluation des apprentissages, en mettant l'accent sur les défis et enjeux qu'ils soulèvent, d'une part, et d'autre part interroge la véracité des informations fournies par des chatbots, tels que ChatGPT, Bing et Consensus, pour l'évaluation des apprentissages avec les outils numériques. 37 articles ont été recensés à travers une recherche dans les bases de données et au moyen des trois robots conversationnels précédemment cités. Les résultats de la présente recherche révèlent le recours par les enseignants à une diversité/pluralité d'outils numériques en soutien à l'évaluation formative, surtout depuis la Covid-19, et un nécessaire alignement entre intentions pédagogiques, curricula et choix d'outils numériques adéquats afin de garantir une évaluation des apprentissages fonctionnelle. De plus, l'évaluation via le numérique charrie des défis autant méthodologiques, techniques, théoriques qu'éthiques. Enfin, les ressources issues des robots conversationnels méritent une vigilance particulière ce qui signifie qu'il faudra probablement un certain temps avant qu'ils puissent être utilisés de manière utile en tant que compléments à la recherche documentaire et à l'évaluation des apprentissages.

MOTS-CLÉS • Évaluation des apprentissages - Outils numériques - Littérature numérique - Robots conversationnels.

ABSTRACT The use of digital tools/instruments brings with it new teaching and learning practices, as well as concerns about the protection of sensitive data. A major challenge is to ensure the validity and reliability of assessment results, by properly calibrating measurement tools and preventing plagiarism. Similarly, the emergence of the use of conversational robots in education raises questions of fairness and authenticity. This article examines digital tools used for learning assessment, focusing on the challenges and issues they raise, on the one hand, and questions, on the other hand, the veracity of information provided by chatbots, such as ChatGPT, Bing and Consensus, for learning assessment with digital tools. 37 articles were identified through a database search and using the three conversational bots mentioned above. The results of the present research reveal the use by teachers of a diversity/plurality of digital tools in support of formative assessment, especially since Covid-19, and a necessary

alignment between pedagogical intentions, curricula and choice of appropriate digital tools to ensure effective assessment of learning. Moreover, digital assessment poses methodological, technical, theoretical and ethical challenges. Finally, resources derived from conversational robots deserve particular vigilance, which means that it will probably be some time before they can be used usefully as complements to documentary research and learning assessment.

KEYWORDS *Learning assessment - Digital tools - Digital literacy - Chatbots*

Problématique

L'évaluation des apprentissages avec des outils/instruments numériques laisse entrevoir de nouvelles possibilités, mais soulève plusieurs enjeux et défis. Elle touche d'abord la question de l'accessibilité des outils/instruments utilisés qui interroge elle-même le niveau de littératie numérique des apprenant·e·s et des enseignant·e·s. Elle occasionne de nouvelles pratiques d'enseignement et d'apprentissage. Elle questionne également la problématique de la protection des données en termes de risques associés à la collecte et à l'utilisation des données sensibles.

Elle convoque enfin des défis liés à la validité et la fiabilité des résultats obtenus. Ce dernier enjeu implique de s'assurer, d'une part, que les outils/instruments de mesure sont correctement calibrés pour rendre compte avec précision de la performance de l'élève et, d'autre part, que le phénomène de plagiat est jugulé le plus possible.

L'émergence, dans la sphère de l'éducation formelle, de robots conversationnels (*chatbots*) tels que *ChatGPT*, *Bing* et *Consensus* soulève notamment des préoccupations majeures en ce qui concerne l'équité au regard de l'initiation des jeunes aux possibilités et limites de ces nouveaux outils/instruments numériques dits intelligents et aussi quant à l'authenticité des travaux scolaires et académiques dans un environnement technologique en constante évolution.

La solution de bloquer l'accès, à l'intérieur des établissements, à ces nouveaux outils/instruments afin de prévenir le plagiat a des limites importantes ; car cela n'évite pas leur emploi lors des devoirs faits à la maison. Par conséquent, il s'agit d'une alternative qui ne serait efficace que dans la perspective d'une stratégie d'évaluation des apprentissages uniquement orientée vers des examens en présentiel. Par ailleurs, des robots conversationnels comme *ChatGPT* et *Bing* sont conçus de manière à être facilement intégrés par les programmeurs dans d'autres sites web. Il est donc quelque peu utopique d'essayer de bloquer l'accès à tous ces sites dans et en dehors de l'établissement scolaire¹.

De plus, les robots conversationnels, programmés à converser sous forme de texte avec un utilisateur et à répondre à ses questions, se multiplient. Cette abondance et cette diversité de choix rendent encore plus ardue toute tentative de contrôle. De fait, l'environnement numérique sera probablement sous peu envahi par ces outils/instruments dits intelligents. Ainsi, il est envisageable que des moteurs de recherche populaires comme *Bing* et *Google* intègrent prochainement des extensions de robots conversationnels. Déjà, des logiciels comme *Sparrow* de DeepMind ou *Bard* de Google viennent concurrencer *ChatGPT*. Par conséquent, la probabilité qu'ils envahissent littéralement l'environnement d'apprentissage, formel sinon informel, des élèves est grande.

Qu'il se produise en classe ou hors classe, un tel « envahissement » est susceptible d'entraîner des usages pédagogiques se traduisant par une hausse des exigences scolaires car les apprenantes et les

¹ Face à une telle situation, les logiciels de détection du plagiat peuvent s'avérer peu fiables puisque diverses stratégies de contournement émergent constamment, soit par le biais d'autres logiciels de réécriture des textes originaux (*Paraphraser.ai*, *Perplexity.ai*, etc.) ou soit par la traduction successive d'un texte en plusieurs langues, afin d'en modifier la structure et le vocabulaire. Ainsi, traquer les cas de plagiat exige actuellement, à l'enseignant, des compétences technologiques nouvelles d'une part, et d'autre part, du temps et de l'énergie supplémentaires dans sa tâche d'évaluation.

apprenants devront davantage faire preuve non seulement d'une maîtrise des contenus mais de capacités de contextualisation et d'interprétation ainsi que de pensée critique. Puisque la véracité du contenu produit par les robots conversationnels est loin encore d'être acquise à 100 %, cet article interroge l'à-propos, voire l'existence avérée, des références fournies par ChatGPT, Bing et Consensus par rapport à la question de l'évaluation des apprentissages avec les outils/instruments numériques. Plus largement, notre but est de contribuer à l'exploration des usages informés et éthiques des outils/instruments numériques ainsi que de leurs enjeux et défis (pédagogiques et éthiques) pour l'éducation formelle.

Nous avons donc entrepris d'apporter des éléments de réponse aux deux questions suivantes :

- Quels sont les outils/instruments, les enjeux et les défis associés à l'évaluation des apprentissages avec le numérique ?
- Comment les robots conversationnels renseignent sur l'état de la question. Quelle est leur véracité... ?

Méthodologie de recension des écrits

Notre recension a porté sur le thème de l'évaluation des apprentissages avec les outils/instruments numériques en présence et à distance. Elle s'est effectuée en trois étapes distinctes, la première d'octobre 2021 à novembre 2021, la seconde de septembre 2022 à novembre 2022, et la dernière de janvier 2023 à avril 2023. Cette recension a couvert la période de publication allant de 2016 à 2022. Pour analyser la véracité de contenu des robots conversationnels retenus, nous effectuons une analyse comparative entre les résultats d'une recherche documentaire, débutée en 2021, sur les usages, enjeux et défis de l'évaluation des apprentissages avec le numérique et les résultats qu'arrivent à fournir Consensus, ChatGPT et Bing – Bard, de Google n'étant pas encore alors disponible – partant des mêmes mots-clés.

Une recherche dans des bases de données

Durant cette étape, la recherche d'articles s'est effectuée dans sept bases de données habituellement utilisées pour la recherche dans le champ de l'éducation que sont EBSCO-Education Source, ERIC, PsycINFO, Google Scholar, Érudit, Cairn et Sofia. L'accès à ces dernières s'est fait via la bibliothèque numérique Sofia de l'Université Laval. Afin de formuler les requêtes dans ces différentes bases de données, les mots-clés ont été organisés en trois regroupements de termes soit : 1-Évaluation numérique, 2- Enseignement secondaire, 3- Outils/instruments numériques. Les requêtes formulées dans ces différentes bases de données sont les suivantes : (*formative and summative evaluation or evaluation method* or evaluation process or digital assessment or évaluation numérique*) ET (*secondary school or enseignement secondaire*) ET (*ICT or digital technolog* or digital technolog* or technologie numérique or TIC*).

Les textes ont été sélectionnés selon qu'ils traitent des méthodes et stratégies, des outils/instruments, des effets et des enjeux liés à une mise en œuvre efficace des diverses formes d'évaluation en contexte d'apprentissage soit en présentiel ou à distance avec des TIC (technologies de l'information et de la communication). Les niveaux d'enseignement primaire et secondaire ont été privilégiés. Étant donné les innovations perpétuelles observées dans les usages pédagogiques

des outils/instruments numériques, nous avons établi une balise temporelle qui couvre les 5 dernières années (soit de 2016 à 2021). Enfin, un dernier palier de sélection a retenu les publications scientifiques ayant été évaluées par des pairs. Au terme de cette recension, 16 articles anglophones et 3 articles francophones, respectant tous les critères d'inclusion susmentionnés, ont été retenus.

Cette première synthèse des connaissances sur le sujet, a été complétée par une seconde au dernier trimestre de 2022. Cette dernière avait pour ambition de colliger les tout récents travaux sur le sujet notamment les considérations éthiques de l'évaluation sous support numérique, peu abordées par les premiers articles recensés, spécifiquement en ce qui concerne l'apport des outils/instruments numériques dans la quête d'authenticité, d'équité et d'éthique dans le processus d'évaluation. Les bases de données initiales ont été maintenues, de même que les mots-clés et les critères d'inclusion. Néanmoins, pour cette seconde phase de recension, seuls les articles publiés en 2022 ont été retenus. Après lecture des résumés, 20 articles ont été présélectionnés pour analyses approfondies dans une grille standardisée. Au final, nous avons 14 études portant sur le sujet qui nous intéresse.

Méthode suivie pour colliger les écrits en provenance de Consensus

L'étape suivante de la recension a voulu expérimenter une recherche documentaire au moyen de *Consensus*. Il s'agit d'un robot conversationnel qui utilise l'intelligence artificielle pour agréger et analyser des milliers d'études scientifiques, incluant des études récentes, disponibles en ligne. Il a également la particularité de pouvoir calculer un pourcentage de consensus (d'où son nom) des auteurs sur une question particulière. Lors de cette recherche documentaire, si les critères d'inclusion des articles ont été maintenus, les spécificités de l'outil de recherche ont éludé le recours aux requêtes tel qu'il s'opère dans les bases de données habituelles.

En effet, Consensus propose plutôt de répondre à une question en compilant et en analysant automatiquement les résultats des études en lien avec les mots-clés identifiés dans la question posée. De plus, la portion de l'article analysé est très souvent le résumé des auteurs. Pour pouvoir explorer les différents aspects de l'évaluation avec le numérique explorés lors des deux premières recensions, plusieurs questions ont été introduites dans la barre de recherche : - *Does digital assessment have a positive effect on learning in secondary school?* - *Does digital learning assessment work?* - *Does digital learning assessment respect equity?*

Méthode suivie pour colliger les écrits en provenance de ChatGPT

ChatGPT est un robot conversationnel qui a la particularité de pouvoir construire des textes par voie prédictive selon des algorithmes statistiques. Les textes peuvent même être longs et rédigés avec en mémoire le contexte de l'interrogation que lui soumet l'utilisateur. En résulte, une synthèse des données présentes dans un corpus qui s'arrête en 2021 et issues de sources multiples et diverses.

Il peut aussi fournir des informations sur des sujets variés et répondre à des questions longues et/ou très détaillées. Pour effectuer une recherche documentaire sur la problématique de l'évaluation avec le numérique, la requête suivante a été introduite : - *Présente tous les articles publiés ces 7 dernières années sur les enjeux, les outils/instruments et les défis de l'évaluation des apprentissages au moyen du numérique en enseignement secondaire. Sois précis avec des références complètes.*

Méthode suivie pour colliger les écrits en provenance de Bing

Bing est le moteur de recherche de Microsoft qui depuis peu intègre un modèle de recherche d'informations basé sur GPT4 alors que ChatGPT est couramment basé sur GPT3.5 – un GPT est un modèle de langage (Generative Pre-Trained Transformer). Bing a la possibilité de compléter des résultats de recherche par des résumés écrits et de proposer des liens vers des sites identifiés pour des recherches approfondies. Il peut également, comme ChatGPT, rédiger des textes selon les indications de l'utilisateur.

Pour effectuer une recherche documentaire sur la problématique de l'évaluation avec le numérique, une requête identique à celle utilisée pour ChatGPT a été introduite : - *Présente tous les articles publiés ces 7 dernières années sur les enjeux, les outils/instruments et les défis de l'évaluation des apprentissages au moyen du numérique en enseignement secondaire. Sois précis avec des références complètes.*

Au total, 37 articles ont été sélectionnés pour la présente recension, 33 en provenance des bases de données et 4 en provenance de Consensus. Les résultats qui suivent sont regroupés sous deux thèmes principaux : 1) les outils/instruments, enjeux et défis associés à l'évaluation avec le numérique et 2) les robots conversationnels, des performances modestes.

Résultats

Outils/instruments, enjeux et défis associés à l'évaluation avec le numérique

Quelques thèmes sont ressortis de la littérature examinée. Sans trop de surprise, le premier touche l'évaluation des apprentissages pendant la pandémie. Suivent ceux relatifs à la diversité des outils/instruments numériques en soutien, à l'adéquation entre intentions pédagogiques, choix d'outils/instruments numériques et évaluation des apprentissages et aux défis de l'évaluation avec le numérique.

Une place conséquente pour l'évaluation des apprentissages avec le numérique pendant la COVID-19

La pandémie de Covid-19 a eu de nombreuses conséquences sur les dispositifs de formation à tous les niveaux d'enseignement. Elle a par ailleurs induit une explosion et une généralisation de l'enseignement à distance devenue une nécessité notamment durant les multiples périodes de confinement. Le basculement d'un enseignement en présentiel, hybride, comodal ou autres, vers un enseignement d'urgence à distance a mis en lumière, d'une part, la problématique de l'ajuste-

ment des outils/instruments, pratiques et dispositifs d'enseignement ; et, d'autre part, celle des stratégies usitées ou préconisées par les professionnels et les chercheurs pour évaluer les connaissances des apprenants.

La littérature révèle que les enseignants ont dû user de stratégies diverses et généralement innovantes pour pouvoir mesurer les performances scolaires de leurs apprenants. Chia et Zhang (2022) colligent dans leur étude, les points de vue de 95 enseignants de mathématiques (niveaux primaire et secondaire) de Hong Kong sur l'évaluation en ligne durant la Covid-19. Il en ressort que ces derniers considèrent l'évaluation en ligne principalement comme une évaluation pour et de l'apprentissage en ligne. De plus, pour ce qui est des stratégies d'évaluation en ligne, 42 % des participants ont utilisé des travaux en classe virtuelle ou des devoirs pour évaluer l'apprentissage des élèves, 16,30 % ont révélé avoir réalisé des tests ou quizz en ligne, 6,52 % ont mentionné avoir posé des questions lors des séances d'enseignement.

Dans le même ordre d'idées, Doukakis et al. (2022), dans un article qualitatif, analysent la manière dont les enseignants en Grèce ont utilisé les techniques d'évaluation lors de l'enseignement en ligne imposé par la Covid-19 dans des écoles secondaires privées en Grèce.

Ils observent d'abord que les enseignants, pour l'évaluation utilisent des outils/instruments de communication (Webex et Microsoft Teams) pour donner des devoirs, des tests oraux pendant la leçon en ligne, la participation à la leçon, des quizz en ligne et la participation au chat privé. Par contre, ils n'ont pas utilisé les outils/instruments numériques à leur plein potentiel car ne disposant pas de temps suffisant pour la conception pédagogique de l'évaluation et ne maîtrisant pas adéquatement les outils/instruments disponibles.

Explorant les effets du passage à l'enseignement à distance d'urgence sur les pratiques d'évaluation durant le confinement dû à la Covid-19, Panadero et al. (2022) ont mené une enquête auprès de 936 professeurs de tous les niveaux d'enseignement en Espagne. Ils observent un changement dans la participation des élèves à l'évaluation, notamment le choix chez les enseignants d'augmenter l'auto-évaluation et l'évaluation par les pairs. Ils observent également, dans certains établissements, une dégradation des normes de notation (par exemple, certains enseignants du primaire ont avoué avoir abaissé les normes de notation pour s'adapter à cette situation d'urgence).

Observant que la conception actuelle de l'évaluation en ligne n'atteint pas toujours les objectifs pédagogiques souhaités dans les environnements d'apprentissages en ligne, notamment à cause d'une incohérence entre les activités d'apprentissage et les méthodes d'évaluation choisies, El Asame et al. (2022) proposent un modèle hybride pour l'évaluation axé sur l'approche basée sur les compétences et l'approche basée sur les objectifs qui, selon eux, permet la conception d'une e-évaluation efficace. Leur modèle s'adosse sur 6 principes clés: la planification méthodique et contextualisée de l'évaluation en ligne, la formulation des compétences et des objectifs clairs, la définition des compétences en fonction des besoins des apprenants, la diversification des activités d'évaluation et des critères de performance, la diversification des activités d'évaluations ayant

la même complexité et enfin, l'alignement entre le niveau taxonomique de la compétence, les objectifs, les critères de performance et les outils/instruments d'évaluation.

En outre, l'évaluation numérique, pour Schmidt et DeSchryver (2022), impose d'abord, pour être efficace, une littératie numérique chez l'apprenant. Schmidt et DeSchryver (2022) proposent un modèle (qu'ils nomment DAppLit) pour assurer la littératie des applications numériques. Le modèle comprend 4 étapes principales : définir des objectifs d'apprentissage, guider les apprenants à travers l'application avec des instructions explicites avant utilisation, proposer des explorations à faible enjeu, et enfin, faire un débriefing pour résoudre les difficultés rencontrées par les élèves. Le but est que la performance de l'apprenant soit tributaire de sa connaissance des contenus d'apprentissage et non pas de la maîtrise de l'application.

Une diversité d'outils/instruments numériques en soutien à l'évaluation formative

Le moissonnage effectué montre une présence relativement discrète de l'évaluation sommative. Par ailleurs, le corpus retenu montre que les enseignants ont recours à des outils/instruments numériques diversifiés pour leur évaluation. On relève une présence notable de l'évaluation formative. Par exemple, les enseignants d'anglais des collèges privés publics en Indonésie ont utilisé de nombreux outils/instruments d'évaluation formative en ligne : Quizizz, Google Forms, Socrative, et Kahoot! (Firdaus et al., 2022). En plus des traditionnels outils/instruments de communication comme Zoom, d'autres enseignants ont opté pour les outils/instruments Webex et Microsoft Teams (Doukakis et al., 2022). Dans le domaine des langues, les applications numériques telles que Novakid, Google Meet, Google Classroom, Whatsapp et Zoom Meeting sont considérées comme efficaces pour améliorer l'enseignement et l'évaluation de manière innovante et vu leur flexibilité (Susanty et al., 2021). Dans le même ordre d'idées, les e-portfolios semblent avoir été utilisés car ils donnent la possibilité de mesurer les progrès d'apprentissage des apprenant·es dans la perspective d'un processus d'évaluation formative au moyen d'une diversité d'activités susceptibles de rendre compte des gains d'apprentissage. (Namaziandost et al., 2020)

Shin et al. (2022) soutiennent que les applications d'intelligence artificielle (IA), utilisées dans un contexte d'évaluation formative informatisée, peuvent améliorer les processus de prise de décision à tous les niveaux de l'éducation. L'étude observe que la plupart des enseignants doivent se fier à leur jugement pour déterminer le calendrier d'administration des tests, ce qui ne prend pas en compte les besoins et les profils de performance des apprenants. Les chercheurs ont donc conçu un modèle de prédiction (qui, selon les auteurs, est susceptible de déterminer automatiquement les progrès de l'élève durant l'année scolaire) afin de détecter le nombre optimal de tests qui lui convient, d'après leur modèle de calcul. Le modèle LSTM (Long short-term memory) proposé serait une aide pertinente dans la prise des décisions éclairées pour rendre les évaluations formatives informatisées plus efficaces et au service de l'apprentissage des élèves.

Correnti et al. (2022), quant à eux, proposent un système automatisé d'évaluation de l'écriture, nommé eRevise, qui d'après les points de vue des apprenants qui ont participé à l'évaluation du produit, possède une rétroaction automatisée qui leur aurait permis une amélioration de leur pro-

duction écrite. Les enseignants ont également indiqué que les messages de rétroaction proposés par l'outil correspondent à leurs objectifs pédagogiques et ce système renforce la rétroaction qu'ils fournissent eux-mêmes aux apprenants durant leur enseignement.

Une nécessité d'adéquation entre intentions pédagogiques, choix d'outils/instruments numériques et évaluation des apprentissages

Il a souvent été constaté que les échelons inférieurs de la taxonomie de Bloom (mémoriser, comprendre, appliquer) étaient davantage travaillés que les échelons supérieurs (analyser, évaluer et créer). Le numérique vient en soutien à l'enseignant soucieux de cette taxonomie lors de la formulation d'intentions pédagogiques. Les possibilités qu'il perçoit se voudront toutefois respectueuses des besoins et des capacités des élèves afin d'atteindre les résultats d'apprentissage visés (Farhat, 2021). Par exemple, avoir recours à un système de cartographie conceptuelle bidirectionnel basé sur l'évaluation par les pairs afin de renforcer la réflexion critique des apprenants et la qualité du retour d'information de leur évaluation (Hwang et Chang, 2020) puisque l'évaluation par les pairs peut alors s'avérer, dans certaines conditions, plus bénéfique que l'évaluation par l'enseignant (Double et al., 2020).

De même, l'enseignant pourrait expérimenter avec le numérique pour fournir de la rétroaction qui soit soutenue, constructive, différenciée et différée. Selon El hage et Nahed (2020), cette rétroaction présente plus d'avantages que la remédiation classique en termes de développement d'un nouveau rapport à l'erreur et d'une nouvelle relation élève-professeur. Lorsque la rétroaction par les pairs est utilisée, les actions de l'enseignant demeurent néanmoins essentielles au processus formatif puisqu'il joue un rôle de médiateur entre la technologie et les apprenant·e·s (Dalby et Swan, 2018).

Le recours à l'évaluation avec le numérique ne donne pas seulement des informations ponctuelles sur les compétences et performances des apprenants. En effet, l'analyse des traces écrites et de l'activité des périphériques (clavier, souris) sont susceptibles de prédire des résultats à un test d'écriture (Tate et Warschauer, 2019). Des éléments du processus d'enseignement-apprentissage qui étaient difficilement évaluables le deviennent par l'analyse des traces.

Une autre possibilité d'évaluation avec le numérique passe par des jeux afin de déterminer si une expérience d'usage d'un jeu virtuel a un effet mesurable (voire prédictif) sur la façon dont les élèves résoudre le problème à l'avenir. Elle est alors moins axée sur l'évaluation des connaissances à proprement parler que sur l'évaluation des processus d'apprentissage. De plus, elle permettrait un gain d'apprentissage notable chez les élèves qui l'utilisent par rapport à ceux qui ne l'utilisent pas (Cutumisu et al., 2019).

De récentes recherches s'orientent vers la réalité augmentée (RA) qui présentent les bénéfices de combiner la réalité avec un monde virtuel, d'être interactive en temps réel et de prendre en charge la visualisation en trois dimensions. Une étude menée sur des apprenants de niveau primaire a notamment montré qu'une évaluation formative basée sur la RA améliore leur motivation et contribue, grâce au feedback rendu en temps réel, à une compréhension plus rapide de la

tâche par rapport à la méthode traditionnelle (Bhagat et al., 2018). Selon Johri (2020), l'explosion numérique a vu le développement d'outils/instruments technopédagogiques qui présentent des gages de sécurité élevés en même temps qu'ils assurent des gains de ressources (humaines, financières, matérielles) et de traitement. Enfin, chacun de ces outils/instruments peut s'arrimer à un type d'activités et à des objectifs d'évaluation bien précis (Noskova et al., 2016).

De fait, l'évaluation avec le numérique a des avantages. Il permet un gain de temps substantiel, une meilleure gestion des effectifs (les dyades d'élèves ou autres petits groupes sont mieux gérables que tout l'ensemble de la classe), une facilitation de la correction, l'identification des apprenants en difficulté, une comparaison et une classification des niveaux et un meilleur suivi des performances (Rehhali et al., 2022). Un autre de ses avantages, est celui de pouvoir maximiser l'efficacité de l'apprentissage particulièrement en jouant un rôle de médiation à l'apprentissage, dans la motivation et l'engagement des élèves à l'égard de cet apprentissage. L'expérimentation, à l'avènement de la Covid-19, d'évaluations alternatives à celle traditionnelle, a pu relever que les premières exhibent une plus grande valeur formative et motivante pour l'élève (Ong et al., 2021) et qu'elles favorisent la différenciation pédagogique (Raulin, 2020).

De plus, l'usage du numérique s'étend désormais à des sphères très spécifiques des habiletés langagières, notamment celles liées au processus d'identification des mots écrits et de la compréhension en lecture (Auphan et al., 2020). Leur étude démontre que l'évaluation informatisée améliore dans certains cas les critères de validité et de fiabilité des tests sur papier, en offrant des moyens de contrôler la présentation des items et l'enregistrement des mesures hors ligne et en ligne pour l'évaluation de la lecture. Dans le champ spécifique de la réalité augmentée, Bhagat et al. (2018) ont montré que le recours aux évaluations formatives basées sur des interfaces virtuelles en 2D ou en 3D, améliore d'une part les performances des élèves, et d'autre part leur motivation à l'apprentissage par rapport aux évaluations formatives traditionnelles.

Cependant, les articles du corpus ne sont pas totalement unanimes sur les effets positifs sur les performances scolaires des évaluations formatives utilisant les outils/instruments numériques. Par exemple, Faber et Visscher (2018) ont montré, au moyen d'une étude quasi-expérimentale, que l'utilisation d'un outil numérique d'évaluation formative comme Snappet n'avait aucun effet significatif sur la réussite en orthographe chez des élèves de 3e année au primaire. Ainsi, considérant ces nombreux avantages, et vu la complexité des compétences requises (autant chez l'enseignant que chez l'apprenant) dans une évaluation au moyen d'outils/instruments numériques, Webb et Ifenthaler (2018) proposent d'associer des évaluations informatisées aux méthodes d'évaluation traditionnelles, afin de rendre le processus d'évaluation des apprentissages plus complet et efficace.

De nombreux défis relatifs à l'évaluation avec le numérique

Un défi important est celui de trouver le bon équilibre (c'est-à-dire la plus grande adaptabilité possible aux caractéristiques individuelles de l'apprenant) entre l'évaluation automatisée et le rôle actif et "constructiviste" de l'apprenant et de l'enseignant (Webb et al., 2018). L'enseignant doit s'y connaître en matière d'évaluation et être capable de faire converger ses pratiques vers des

modèles qui prennent en compte des spécificités d'apprentissage. Par exemple, l'évaluation numérique multimodale (qui explore un éventail de ressources sémiotiques telles que l'image, l'écriture, le geste, le regard, la parole ou la posture) offre des répertoires plus riches de modalités disponibles pour les apprenants et les enseignants (Fjortoft, 2020), notamment dans le cadre de l'évaluation en anglais (Flanagan et Hall, 2018). Pour certains, l'évaluation qui tire profit du numérique est considérée comme plus authentique, plus contextuelle et plus proche des scénarios du monde (DiCerbo, 2020), sous la condition qu'elle s'attache au respect des variables de diversité.

Dans le cas spécifique d'une évaluation à distance d'urgence, une étude menée auprès de 17 enseignants du secondaire de l'Ontario au Canada révèle qu'elle induit 3 contraintes principales : une fréquence accrue des rétroactions pour soutenir l'apprentissage des élèves à la maison, le passage à une approche authentique basée sur des tâches non plagiantes et le retour à l'évaluation formative comme facteur important du soutien à l'apprentissage (Cooper et al., 2022). De plus, il se pose un autre défi d'authenticité des évaluations en ligne. Par exemple, DeCoito et Estaiteyeh, (2022) ont effectué une étude mixte (qualitative et quantitative) dans une province du Canada auprès de 75 enseignants de sciences de la 1^{ère} à la 12^{ème} année sur les pratiques d'évaluation durant la Covid et leur efficacité. Les participants ont jugé les techniques d'évaluation non authentiques et peu efficaces. En effet, ils ont estimé qu'elles ne reflétaient pas fidèlement les attentes scolaires en termes de mesure de la compréhension des notions et de la maîtrise des compétences. L'inauthenticité émanait également des problèmes d'équité et d'accessibilité à la technologie.

Allant dans la même logique, Nisbet et Shaw (2022) estiment, dans le contexte imposé par la Covid-19, qu'il existe trois grands défis à une évaluation en ligne équitable : un défi théorique, un défi lié à la réflexion sur la justice sociale et un défi lié à la stratégie d'attribution des notes. Plus spécifiquement, au niveau théorique, les auteurs soutiennent que la validité et la large accessibilité d'un test sont requises dans la quête d'équité, mais que l'équivalence l'est un peu moins. En outre, la justice sociale voudrait que la perte différentielle d'apprentissage des élèves à la suite de la pandémie soit prise en compte dans le processus de sélection pour l'entrée au collège et à l'université.

Du Plessis (2022), quant à lui, propose des réflexions personnelles et subjectives sur l'enseignement, l'apprentissage et l'évaluation en ligne durant et après la Covid en Afrique du Sud. D'après lui, la pandémie a entraîné un changement important dans les pratiques d'évaluation des connaissances. L'évaluation désormais exige l'engagement des apprenants dans des réflexions critiques en ligne, des présentations de groupes en ligne (notées individuellement et en groupe). En outre, d'après le chercheur, lorsque l'évaluation suit le modèle des questions à choix multiples (QCM) cela induit une charge de travail supplémentaire en termes de temps alloué à la préparation et à la conception des questions.

Les obstacles énumérés par Rehhali et al. (2022) concernent les défis techniques liés au manque de compétence technologique des enseignants, les contraintes didactiques liées à la conception

des situations d'évaluation et les contraintes organisationnelles liées à la programmation des évaluations en ligne. De ce fait, les enseignants doivent améliorer leur culture informatique et numérique pour soutenir l'application efficace des évaluations en ligne (Firdaus et al., 2022).

Deux conceptions éthiques de l'évaluation ressortent particulièrement du discours de certains apprenants du primaire et du secondaire, l'une liée à l'égalité et l'autre liée à l'équité (Murillo et Hidalgo, 2017). De leur point de vue une évaluation sous le sceau de l'égalité est conçue comme impartiale, quantitative et explicite, tandis qu'une évaluation équitable est définie comme contextualisée, qualitative et procédurale. Les auteurs partent du postulat que si les élèves considèrent les évaluations comme injustes, ces dernières peuvent avoir sur leur apprentissage un effet positif moindre que s'ils pensent le contraire. Enfin, et ce d'un point de vue général, les concepteurs d'évaluations devraient certes se conformer à des théories d'apprentissage précises, mais également concevoir des tâches diversifiées qui saisissent mieux l'éventail des idées et des expériences des élèves (Kang et Furtak, 2021).

En somme, les usages du numérique offrent de nouvelles possibilités pédagogiques, mais présentent des défis surtout pour des enseignants peu expérimentés en matière à la fois d'intégration et d'évaluation avec des outils/instruments numériques. Toutefois, il paraît primordial d'adosser les pratiques d'usages du numérique à des modèles théoriques solides.

Les robots conversationnels, des performances modestes

Le recours aux robots conversationnels a mis en évidence un certain écart avec la méthodologie de recension via les bases de données, des biais méthodologiques manifestes et des résultats quelque peu mitigés en termes de contenu obtenu.

Consensus

Pour rappel, trois requêtes ont été introduites dans la barre de recherche de Consensus : - *Does digital assessment have a positive effect on learning in secondary school?* - *Does digital learning assessment work?* - *Does digital learning assessment respect equity?*

Le robot a calculé pour ces questions des pourcentages de consensus respectifs de 100 % pour les deux premières, ce qui est une information utile en soi malgré le fait qu'il s'en est tenu aux cinq premiers articles jugés les plus pertinents. Toutefois, il n'a pas été en mesure d'en faire de même pour la dernière, jugeant les résultats des articles pas assez pertinents. Ce robot ne peut donc fournir, du moins pour le moment, une analyse profonde et précise des articles auxquels il a accès puisqu'il s'en tient qu'aux résumés de 5 articles à la fois. De plus, sa capacité d'analyse du consensus n'affiche que les 20 premiers articles jugés les plus pertinents. Pour établir la pertinence des articles, *Consensus* suit un processus précis. Tout d'abord, il effectue une recherche basée sur les mots-clés de la requête dans sa base de données d'articles scientifiques en ligne. Puis, le modèle d'intelligence artificielle évalue les relations entre la requête et les résultats possibles et établit un score de pertinence pour chaque article. Enfin, ce score de pertinence est combiné à d'autres métadonnées comme le nombre de citations et la date de publication afin de générer un top 20 des résultats.

Cependant, à la différence des bases de données habituelles, Consensus n'affiche pas une estimation du nombre total d'articles trouvés ou du nombre de pages des résultats, comme le ferait Google Scholar par exemple. Par conséquent, la présélection s'est arrêtée lorsque les articles proposés semblaient de moins en moins pertinents. De la somme d'articles présélectionnés, 5 (dont un avait été au préalable trouvé grâce aux bases de données) respectaient les critères d'inclusion prédéfinis.

ChatGPT

Le robot a d'emblée admis ne pas être en mesure de présenter l'entièreté de la recherche scientifique sur le sujet et s'est limité à 8 articles. Il a fallu le relancer à de nombreuses reprises afin qu'il présente des articles supplémentaires. Après 5 requêtes similaires, ChatGPT a rendu un total de 37 articles, dont 20 étaient erronés (parce qu'inexistants), 15 étaient exacts mais en dehors des balises temporelles de publication et 2 étaient des doublons.

La vérification de l'existence réelle des articles s'est effectuée par une recherche des auteurs et des titres sur Google, combinée à une recherche dans les revues scientifiques identifiées par le robot dans les références complètes. Par ailleurs, une lecture des 15 articles exacts a mis en lumière le fait qu'ils n'auraient, de toute façon, pas été pertinents pour la présente recension. En effet, 7 articles ne traitent pas de l'évaluation des apprentissages (mais plutôt de l'usage des médias sociaux ou des TIC en enseignement ou encore des stratégies/méthodes d'enseignement), 2 articles explorent effectivement la thématique de l'évaluation mais uniquement du point de vue de la différence formelle entre l'évaluation formative et celle sommative (sans recours aux outils/instruments numériques), seuls 6 articles cadrent avec le thème de l'évaluation avec les outils/instruments numériques.

S'agissant des contenus desdits articles, ChatGPT annonce qu'ils traitent de certains thèmes qui pourtant ne sont pas indiqués dans la requête de recherche (l'impact de l'apprentissage mobile sur la motivation et la réussite des élèves, la littératie numérique, l'utilisation de la technologie pour soutenir l'apprentissage autonome) et d'autres thèmes plus pertinents (l'utilisation de la technologie pour l'évaluation en STEM, l'évaluation formative basée sur les jeux, les portfolios numériques pour l'évaluation en mathématiques, l'utilisation de l'évaluation formative numérique en langue anglaise, l'évaluation des apprentissages en STEM à travers des projets basés sur la technologie...) mais qui pour la plupart concernent des articles erronés. Finalement, aucun article issu de ChatGPT n'aura été retenu pour la présente recension. En outre, dans le but d'analyser les capacités du robot à synthétiser une masse importante d'informations et, d'autre part, de juger de leur véracité, la première requête a été reformulée en ces termes : - Que disent les études et la recherche sur les outils/instruments, les enjeux et les défis associés à l'évaluation au moyen des outils/instruments numériques en enseignement secondaire. Sois précis dans les sources, les travaux, les auteurs et les résultats des recherches que tu cites.

S'agissant du contenu, le robot aura effectivement présenté des outils/instruments, des enjeux et des défis similaires voire complémentaires à ceux observés dans la revue littérature effectuée plus haut. Pour ce qui est des outils/instruments numériques, ChatGPT les a organisés en trois

grandes catégories : les plateformes de création des évaluations en ligne (Kahoot, Quizlet, Google Forms et Edpuzzle), les outils/instruments de correction automatique (Turnitin, Grammarly) et les outils/instruments de rétroaction (Google Docs, Bubble). Si l'on peut s'interroger sur la présence dans cette liste de Bubble, qui est une plateforme permettant de créer des applications sans lignes de codes, les outils/instruments proposés par le robot conversationnel n'avaient pas été présentés dans les articles sélectionnés.

De plus, il aura proposé des enjeux différents comme la rapidité des rétroactions, l'augmentation de l'objectivité dans la correction ou encore la place de la surveillance humaine afin de pallier les erreurs éventuelles des outils/instruments utilisés. Enfin, il aura mis en exergue des défis techniques (en lien avec la disponibilité des outils/instruments, leur accessibilité, surtout pour des élèves ayant des besoins particuliers et leur calibrage), pédagogiques (équivalence en termes de qualité aux évaluations traditionnelles) et éthiques (confidentialité des données, protection de la vie privée des élèves, sécurisation des données collectées, validité et fidélité des instruments d'évaluation, fiabilité des résultats des évaluations).

Au demeurant, le recours à ChatGPT aura permis d'obtenir des informations complémentaires sur le thème à l'étude. Cependant, si les informations en elles-mêmes sont pertinentes, les sources proposées sont toutes erronées, soit parce qu'elles renvoient à des articles inexistantes, soit parce que le contenu proposé par ChatGPT est faux (nom erroné de revue par exemple). Au final, le robot conversationnel représente sans doute un atout non négligeable pour acquérir des connaissances globales, mais est plus limité dans le cadre d'une recherche scientifique.

Bing

Bing a répondu n'avoir trouvé aucun article correspondant à notre demande. Il a également proposé trois liens vers les sites web de 2 universités canadiennes (université Laval et université du Québec à Chicoutimi) et du Ministère de l'éducation et de l'enseignement supérieur (qui est pourtant, à l'heure actuelle, éclaté en deux ministères distincts au Québec). De plus, les sites web indiqués ne proposaient pas d'articles, mais des ressources pédagogiques générales sur l'évaluation des apprentissages aux niveaux secondaire et universitaire. En résumé, aucun article additionnel provenant de Bing et aucune information supplémentaire en lien avec le thème de recherche n'ont été trouvés.

Discussion et perspectives

À partir de la présente recension des écrits, qui explore les usages en évaluation des apprentissages des outils/instruments numériques ainsi que leurs enjeux et défis pour l'éducation formelle selon la méthodologie conventionnelle, et qui interroge aussi l'à-propos des références fournies par Consensus, ChatGPT et Bing, nous dégageons quelques constats.

Le premier constat est que l'apport des outils/instruments numériques est encore fort peu documenté compte tenu, d'une part, de leurs fonctionnalités croissantes et, d'autre part, des enjeux et défis relatifs à leurs usages, voire des enjeux et défis relatifs à l'accompagnement des ensei-

gnants (Nadeau-Tremblay et al., 2022). Qui voudra se passer dans quelques années des outils/instruments informatiques permettant de faire de l'évaluation, notamment dans le domaine de l'évaluation formative, y compris d'outils/instruments non conçus pour faire de l'évaluation mais permettant d'échanger à distance, de manière synchrone ou asynchrone ?

Le deuxième constat est que les inévitables systèmes de quizz, modernisés pour traiter des données multimédias sont toujours là au service de l'évaluation sommative et on ne voit pas comment ils pourraient disparaître à court terme, cela même alors que les robots conversationnels se développent. S'ils ont des inconvénients bien documentés comme le fait d'enseigner uniquement pour faire réussir au test, le bachotage voire l'incitation à la triche quand les tests sont à fort enjeu (Amrein-Beardsley & al, 2010), ils sont bien implantés, suppriment le travail de correction et produisent des résultats quantifiés fort appréciés des administrations. De plus, certains sont soigneusement conçus et peuvent être utilisés dans une perspective formative (voir par exemple Lavigne 2022 à partir du système PIX d'évaluation des compétences dans le domaine du numérique).

Le troisième constat est que, ce qui compte, finalement, c'est l'alignement de ces outils/instruments avec les curricula, les intentions et les pratiques pédagogiques, et, bien sûr, les défis éthiques et pratiques soulevés par l'évaluation informatisée. Bien avant ChatGPT, un de ses précurseurs, LSA (*Latent Semantic Analysis*) a été envisagé pour cela (Lemaire et Dessus, 2015). Ces auteurs utilisaient l'argument que d'après eux :

LSA rend opératoires les théories associationnistes qui postulent depuis Aristote un fonctionnement cognitif fondé sur l'association d'éléments irréductibles. Aristote a proposé quatre formes d'association : la similarité, la différence, la contiguïté temporelle et la contiguïté spatiale (p. 8).

Teplovs et Scardamalia (2007) ont fait des essais intéressants avec la LSA qui se sont notamment traduits en visualisations des interactions écrites d'élèves. Mais l'idée de l'évaluation sans responsabilité d'une personne humaine pose tellement de questions qu'elle est difficilement acceptable pour des enseignants. Tant que ces derniers conservent leur agentivité actuelle, il est peu probable que de telles méthodes puissent se banaliser à court terme.

Enfin, avant d'évaluer il faut bien faire apprendre, aux enseignant·e·s et aux élèves, l'usage des outils/instruments numériques et les enjeux éthiques qui y sont associés. On est là sans doute dans un sous-domaine de ce qu'il est convenu d'appeler les humanités numériques, où la vigilance des professionnels de l'éducation est nécessaire.

Le quatrième et dernier constat concerne l'usage des robots conversationnels dont la présence dans la recherche documentaire est actuellement modeste, voire contre-productive. Le recours à Bing a été infructueux, tandis que dans l'ensemble des études proposées par ChatGPT, plus de la moitié des articles s'avéraient inexistantes, et les autres ne respectaient pas le critère lié à la date

de publication (et parfois aussi celui lié au thème de recherche). Seul Consensus aura permis de compléter la recension effectuée dans les bases de données, avec des références toujours exactes.

Ces systèmes nécessitent donc une vigilance particulière. Il va donc se passer probablement du temps avant qu'ils puissent servir de compléments pertinents autant à la recherche documentaire qu'à l'évaluation des apprentissages.

Références bibliographiques

- Amin, R. & Rahayu, S. (2022). Education evidence approach: Type validity of Computer Based Test (CBT) in formative and summative assessment for vocational high schools. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, 5(1), 3323-3329. <https://bircu-journal.com/index.php/birci/article/view/3999>
- Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education policy analysis archives*, 18, 32. <http://epaa.asu.edu/ojs/article/view/714>
- Auphan, P., Ecalle, J., & Magnan, A. (2020). The High Potential of Computer-Based Reading Assessment. *Canadian Journal of Learning and Technology*, 46 (1). <https://doi.org/10.21432/cjlt27847>
- Bhagat, K. K., Wei-Kai Liou, W-K, Spector, J. M. & Chang, C-Y. (2018). To use augmented reality or not in formative assessment: a comparative study. *Interactive Learning Environments*, 1-11. <https://doi.org/10.1080/10494820.2018.1489857>
- Chia, H. M. & Zhang, Q. (2022). Assessment of/for/as online learning: Mathematics teachers' views on online assessment during the COVID-19 pandemic. Twelfth Congress of the European Society for Research in Mathematics Education (CERME12). <https://hal.archives-ouvertes.fr/hal-03753411>
- Cooper, A., DeLuca, C., Holden, M. & MacGregor, S. (2022). Emergency assessment: rethinking classroom practices and priorities amid remote teaching. *Assessment in Education: Principles, Policy & Practice*, 1-21. <https://www.tandfonline.com/doi/full/10.1080/0969594X.2022.2069084>
- Correnti, R., Matsumura, L. C., Wang, E. L. & Litman, D. (2022). Building a validity argument for an automated writing evaluation system (eRevise) as a formative assessment. *Computer and Education Open*, 3, 1-15. <https://doi.org/10.1016/j.caeo.2022.100084>
- Cutumisu, M., Chin, D., & Schwartz, D., L. (2019). A Digital Game-Based Assessment of Middle-School and College Students' Choices to Seek Critical Feedback and to Revise. *British Journal of Educational Technology*, 50(6), 2977-3003. <https://doi.org/10.1111/bjet.12796>
- Dalby, D. & Swan, M. (2019). Using Digital Technology to Enhance Formative Assessment in Mathematics Classrooms. *British Journal of Educational Technology*, 50(2), 832-845. <https://doi.org/10.1111/bjet.12606>
- DeCoito, I. & Estaiteyeh, M. (2022). Online teaching during the COVID-19 pandemic: exploring science/STEM teachers' curriculum and assessment practices in Canada. *Disciplinary and*

Interdisciplinary Science Education Research, 4(1), 1-18. <https://doi.org/10.1186/s43031-022-00048-z>

- DiCerbo, K. (2020). Assessment for Learning with Diverse Learners in a Digital World. *Educational Measurement: Issues and Practice*, 39(3), 90-93. <https://doi.org/10.1111/emip.12374>
- Double, K.S., McGrane, J.A. & Hopfenbeck, T.N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educ Psychol Rev*, 32, 481–509. <https://doi.org/10.21432/cjlt27847>
- Doukakis, S., Niari, M., Alexopoulos, E. & Sfyris, P. (2022). Online Learning, Students' Assessment and Educational Neuroscience. Dans Auer, M. E. & Tsiatsos, T. (Eds). *New Realities, Mobile Systems and Applications*, 411. (p. 71-82). https://link.springer.com/10.1007/978-3-030-96296-8_7
- du Plessis, A. (2022). A personal reflection on Covid-19 online teaching, learning, and assessment with suggestion for research. *Journal of Baltic Science Education*, 21, 536-544. <https://doi.org/10.33225/jbse/22.21.536>
- El Asame, M., Wakrim, M. & battou, A. (2022). Designing e-assessment activities appropriate to learner's competency levels: Hybrid pedagogical framework and authoring tool. *Education and Information Technologies*, 27(2), 2543-2567. <https://doi.org/10.1007/s10639-021-10607-y>
- El Hage, F., & Nahed, R. (2020). Apprentissage par Exploitation de l'Erreur et à Distance (AEED). Pour une évaluation formative et un feedback interactif et digital. *Évaluer. Journal international de recherche en éducation et formation*, 1, 131-142.
- Faber, J. M & Visscher, A. J. (2018). The effects of a digital formative assessment tool on spelling achievement: Results of a randomized experiment. *Computers & Education*, 122, 1-8. <https://doi.org/10.1016/j.compedu.2018.03.008>
- Farhat, N., H. (2021). Use of digital assessment. How to utilize digital Bloom to accommodate online learning and assessment? *Asian journal of education and training*, 7 (1), 30-35.
- Firdaus, M., Prastikawati, E., & Wiyaka, W. (2022). Online Formative Assessments in English Teaching and Learning. *SALÉE: Study of Applied Linguistics and English Education*, 3(1), 23-34.
- Fjortoft, H. (2020). Multimodal digital classroom assessments. *Computer & Education*, 152. <https://doi.org/10.1016/j.compedu.2020.103892>
- Flanagan, E., & Hall, T. (2017). Digital Ensemble: The ENaCT Design-Based Research Framework for Technology-Enhanced Embodied Assessment in English Education. *English in Education*, 51 (1), 76-99. <https://doi.org/10.1111/eie.12136>
- Hwang, G.-J. (2021). Facilitating Knowledge Construction in Mobile Learning Contexts: A Bi-Directional Peer-Assessment Approach. *British Journal of Educational Technology*, 52 (1), 337-357. <https://doi.org/10.1111/bjet.13001>
- Johri, A. (2020). IT Tools and Technologies for Examination & Assessment. International conference on teaching, assessment and learning in the digital age. Conference proceedings, 77-89.

- Kang, H., & Furtak, E. M. (2021). Learning theory, classroom assessment, and equity. *Educational Measurement: Issues and Practice*. Advance online publication. <https://doi.org/10.1111/emip.12423>
- Laferrère, T., Baron, G.-L., Allaire, S., Nadeau-Tremblay, S., Tremblay, M., & Beaudoin, F. B. M.-C. N. et J. (2022). Numérique, éducation et forme scolaire : Enjeux d'équité. *Diversité. Revue d'actualité et de réflexion sur l'action éducative*, 200. <https://doi.org/10.35562/diversite.1625>
- Lavigne, M. (2022). Les enjeux des plateformes numériques d'enseignement gamifiées : Enquête d'utilisation de la plateforme Pix. *Médiations et médiatisations*, 11, 94116. <https://revue-mediations.teluq.ca/index.php/Distances/article/download/279/249>
- Lemaire, B., & Dessus, P. (2003). Modèles cognitifs issus de l'Analyse de la Sémantique Latente. In *Cognito-Cahiers Romains de Sciences Cognitives*, 1(1), 5574. <https://hal.univ-grenoble-alpes.fr/hal-01222929/document>.
- Murillo, F., J. & Hidalgo, N. (2017). Students' conceptions about a fair assessment of their learning. *Studies in Educational Evaluation*, 53, 10-16. <https://doi.org/10.1016/j.stueduc.2017.01.001>
- Nadeau-Tremblay, S., Tremblay, M., Laferrère, T., & Allaire, S. (2022). Les enjeux et défis d'accompagnement d'enseignantes et d'enseignants dans l'évaluation des apprentissages à l'aide de technologies collaboratives au primaire et au secondaire. *Médiations et médiatisations*, 9, 7-27. <https://revue-mediations.teluq.ca/index.php/Distances/article/view/249/208>
- Namaziandost, E., Alekasir, S., Mohammed Sawalmeh, M. H. & Zaini Miftah, M. (2020). Investigating the Iranian EFL learners' attitudes towards the implementation of e-portfolios in English learning and assessment. *Cogent Education*, 7(1), 1856764. <https://doi.org/10.1080/2331186X.2020.1856764>
- Nisbet, I. & Shaw, S. (2022). Fair high-stakes assessment in the long shadow of Covid-19. *Assessment in Education: Principles, Policy & Practice*, 1-16. <https://doi.org/10.1080/0969594X.2022.2067834>
- Noskova, T., Pavlova, T. & Yakovleva, O. (2016). Approach to Selecting ICT Tools for Formative Assessment. 11 th International Scientific Conference on Distance Learning in Applied Informatics, 200-214.
- Ong, M., H., A., Yasin, N., M., & Ibrahim, N., S. (2021). Immersive Experience during Covid-19: The Mediator Role of Alternative Assessment in Online Learning Environment. *International Journal of Interactive mobile Technologies*, 15 (18), 16-32. <https://doi.org/10.3991/ijim.v15i18.24541>
- Raulin, D. (2020). L'évaluation des acquis des élèves, en période de confinement. Évaluer. *Journal international de recherche en éducation et formation*. Numéro Hors-série 1, 231-236.
- Panadero, E., Fraile, J., Pinedo, L., Rodríguez-Hernández, C. & Díez, F. (2022). Changes in classroom assessment practices during emergency remote teaching due to COVID-19. *Assessment in Education: Principles, Policy & Practice*, 29(3), 361-382. <https://doi.org/10.1080/0969594X.2022.2067123>

- Rehhali, M., Mazouak, A. & Belaouad, S. (2022). The digital assessment of learning: Current situation and perspectives: Case of teachers of life and earth Sciences. *Journal of Information Technology Management*, 14(3), 65-78. <https://doi.org/10.22059/jitm.2022.87534>
- Shin, J., 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>
- Schmidt, L. & DeSchryver, M. (2022). The role of digital application literacy in online assessment. *Journal of Educational Technology Systems*, 50(3), 356-378. <https://doi.org/10.1177/00472395211052644>
- Susanty, L., Hartati, Z., Sholihin, R., Syahid, A., & Liriwati, F. Y. (2021). Why English teaching truth on digital trends as an effort for effective learning and evaluation: opportunities and challenges: analysis of teaching English. *Linguistics and Culture Review*, 5(1), 303-316. <https://doi.org/10.37028/lingcure.v5nS1.1401>
- Tate, T., P. (2019). Keypresses and Mouse Clicks: Analysis of the First National Computer-Based Writing Assessment. *Technology, Knowledge and learning*, 24(4),523-543. <https://doi.org/10.1007/s10758-019-09412-x>
- Teplovs, C., & Scardamalia, M. (2007). *Visualizations for knowledge building assessment*. Paper presented at the Agile Viz workshop, CSCL.
- Webb, M., Ifenthaler, D. (2018). Assessment as, for, and of Twenty-First Century Learning Using Information Technology: An Overview. In: Voogt, J., Knezek, G., Christensen, R., Lai, KW. (eds) *Second Handbook of Information Technology in Primary and Secondary Education*. Springer International Handbooks of Education. Springer, Cham. https://doi.org/10.1007/978-3-319-71054-9_37
- Webb, M., Prasse, D., Phillips, M., Kadijevich, D. M., Angeli, C., Striker, A., Carvalho, A. A., Andresen, B. B. Dobozy, E. & Laugesen, H. (2018). Challenges for IT-Enabled Formative Assessment of Complex 21st Century Skills. *Technology, Knowledge and learning*, 23, 441-456. <https://doi.org/10.1007/s10758-018-9379-7>

ANNEXE : Identification des articles par source de provenance

Articles	Bases de données				Consensus
	EBSCO	PsyInfo	Google Scholar	Sofia	
Amin et Rahayu (2022)			X		
Auphan <i>et al.</i> (2020)	X				
Bhagat <i>et al.</i> (2018)	X				X
Chia et Zhang (2022)			X		
Cooper <i>et al.</i> (2022)			X		
Correnti <i>et al.</i> (2022)			X		
Cutumisu <i>et al.</i> (2019)	X				
Dalby et Swan (2019)	X				
DeCoito et Estaiteyeh (2022)			X		

DiCerbo (2020)				X	
Double <i>et al.</i> (2020)				X	
Doukakis (2022)			X		
du Plessis (2022)			X		
El Asame <i>et al.</i> (2022)			X		
El Hage et Nahed (2020)			X		
Faber et Visscher (2018)					X
Farhat (2021)	X				
Firdaus <i>et al.</i> (2022)			X		
Fjortoft (2020)			X		
Flanagan et Hall (2017)	X				
Hwang (2021)	X				
Johri (2020)			X		
Kang et Furtak (2021)		X			
Murillo et Hidalgo (2017)	X				
Namaziandost <i>et al.</i> (2020)					X
Nisbet et Shaw (2022)			X		
Noskova <i>et al.</i> (2016)			X		
Ong <i>et al.</i> (2021)	X				
Raulin, D. (2020)			X		
Panadero <i>et al.</i> (2022)			X		
Rehhali <i>et al.</i> (2022)			X		
Shin <i>et al.</i> (2022)			X		
Schmidt et DeSchryver (2022)			X		
Susanty <i>et al.</i> (2021)					X
Tate (2019)	X				
Webb et Ifenthaler (2018)					X
Webb <i>et al.</i> (2018)			X		