

Chapitre 21. Du testing adaptatif à l'évaluation intelligente

40 ans de tissage entre le numérique et l'évaluation au sein
de la Communauté ADMEE

Pierre-François Coen

I. Introduction

Le lien entre technologie et apprentissage n'est pas récent. Depuis les machines à enseigner chères à Skinner aux applications actuelles de l'IA en passant par l'apprentissage adaptatif ou la réalité virtuelle, l'espoir d'améliorer les performances des élèves, de comprendre leur manière d'apprendre ou encore d'individualiser l'enseignement a toujours été présent durant ces dernières décennies. Ainsi, que ce soit dans sa fonction formatrice ou certificative, l'évaluation est intimement liée à l'apprentissage. Il apparaît donc comme normal et évident que l'évaluation soit elle aussi très liée à la technologie. À la recherche de plus-value qualitative ou plus simplement pour gagner du temps, l'idée de technologiser l'évaluation est là depuis longtemps et a fait, elle aussi, un chemin impressionnant.

L'enjeu de ce chapitre est de voir comment la Communauté ADMEE a traité cette question. Nous allons donc retourner en arrière et faire un pas de côté pour prendre un peu de recul et voir comment le couple numérique - évaluation a évolué ces quarante dernières années. Pour guider cette réflexion, nous allons nous appuyer essentiellement sur les travaux effectués au sein de la communauté de l'ADMEE en nous fondant principalement sur les écrits publiés dans les revues *Mesure et évaluation en éducation* (MEE) et *Journal international de recherche en éducation et formation* (E-JIREF). Dans cet exercice, nous ne revendiquons évidemment pas une analyse exhaustive de toutes les traces laissées durant toutes ces années (nous aurions pu aussi examiner tous les programmes des colloques de l'ADMEE), mais après un examen minutieux d'une cinquantaine d'écrits, il nous semble pertinent de donner quelques orientations susceptibles d'enrichir les réflexions actuelles. Précisons encore que nous nous sommes centrés sur des écrits significatifs à nos yeux qui prenaient pour objets le numérique. Nous avons délibérément laissé de côté ceux qui touchaient au numérique de manière indirecte, par exemple par le biais de méthodologies intégrant des outils numériques comme la vidéo.

Nous commencerons par donner quelques points de repères historiques et terminologiques. Puis, le chapitre se découpera en cinq sections qui croiseront à la fois des thématiques emblématiques et des balises temporelles. Nous terminerons par quelques constats conclusifs.

II. Quelques points de repère

Disons d'emblée que le couple évaluation - informatique est relativement peu présent dans la communauté ADMEE. En 1991, les rédacteur·trice·s de la revue MEE présentent une bibliographie thématique des 14 premiers volumes de la revue (de 1978 à 1991) et dénombrent relativement peu d'articles consacrés à ce thème. Comparativement aux autres thèmes, cette catégorie reste très marginale. En observant les publications de la revue MEE sur une plus longue période (voir figure n° 1), on peut dire qu'en moyenne, un article traitant de l'informatique/l'ordinateur/le numérique et l'évaluation est publié à peu près tous les deux trois ans. La cadence est à peine supérieure au début des années 90 qui coïncident avec la

démocratisation des P.C. En 1990, un premier symposium autour du rôle des ordinateurs et de l'évaluation est organisé et aborde principalement des questions autour du testing adaptatif et diagnostique. Au fil des ans, les articles traitant des thématiques de validation des tests, de banque d'items, puis de numérique et dispositifs d'apprentissage (plateforme e-learning) s'égrènent encore timidement. En 2019, un symposium au colloque de 2018, conduit à la publication d'un numéro d'E-JIREF consacrée au numérique (Coen et al. 2019) et, une année plus tard, un opus spécial consacré à l'évaluation durant la pandémie du COVID permet le partage de nombreuses pratiques d'évaluation en particulier dans des contextes d'enseignement à distance.

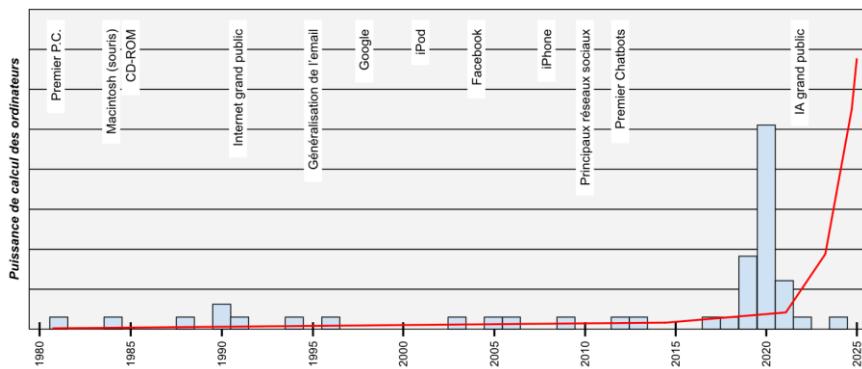


Figure 1 : Nombre de publications consacrées au numérique et à l'évaluation avec la courbe explicitant la puissance des ordinateurs ces 50 dernières années

Lorsqu'on parcourt les textes, on peut voir que différents termes ont été utilisés pour désigner ce qu'on appelle aujourd'hui les technologies numériques : informatique, ordinateur (micro-ordinateur) et évaluation, système informatisé, gestion informatisée de l'enseignement et des apprentissages, évaluation assistée par ordinateur, logiciels d'évaluation, NTIC (nouvelles technologies de l'information et de la communication) ou TIC (technologies de l'information et de la communication) et évaluation, apprentissage adaptatif, évaluation et intelligence artificielle, etc. Cette diversité montre une grande mouvance du champ et une certaine instabilité. Elle n'est pas surprenante et résulte sans doute de l'évolution très rapide des technologies dans la société durant ces cinquante dernières années. La figure n° 1 montre également le développement important (en

particulier ces 5 dernières années) de la puissance de calcul des ordinateurs. Selon certains analystes, cette puissance a doublé quasiment tous les deux ans depuis 1960 et depuis l'avènement de l'IA, les choses se sont encore grandement accélérées (Connes, 2022) et touchent de plein fouet les pratiques d'évaluation ; nous y reviendrons.

Examinons maintenant quelques thématiques saillantes issues de la communauté ADMEE. Nous avons pris le parti de faire a priori une présentation chronologique des thématiques, même si cette règle ne sera pas toujours strictement suivie.

III. Les premiers pas du testing adaptatif

Le premier article explicitant clairement le rôle de l'ordinateur dans les pratiques d'évaluation est celui de Seguin (1984) intitulé *L'utilisation des micro-ordinateurs pour l'évaluation des apprentissages : quelques perspectives d'avenir*. L'auteur commence par dire que si l'ordinateur est utilisé comme support pour les apprentissages, il doit l'être également pour les évaluer. Il poursuit sa réflexion en évoquant une nouveauté importante pour l'époque : le testing adaptatif. Il note ainsi tout l'intérêt d'utiliser l'ordinateur pour constituer des banques d'items dont la passation serait grandement facilitée via l'ordinateur. Le paramétrage des items en termes de difficulté ainsi que « la conservation et l'analyse des résultats obtenus pour vérifier la validité de chaque item par une technique d'analyse scientifiquement éprouvée et reconnue » (p. 55) serait un atout indéniable de même que la possibilité de réduire la charge administrative (impression des notes, correction d'examens, mise à jour des dossiers cumulatifs). Pour lui, l'usage de l'ordinateur pourrait être fait dans une « évaluation formative par étapes itératives » (p. 60) au cours de laquelle l'élève répond à des questions et les réponses qu'il donne permettent de situer son niveau. La complexité de la démarche serait grandement facilitée par l'utilisation de l'ordinateur et conduirait petit à petit à ce que l'auteur appelle le testing adaptatif. Il termine par un plaidoyer intéressant sur « la gestion informatisée de l'évaluation pédagogique soutenant la mise à disposition de logiciel pour les enseignants pour leur faciliter l'accomplissement des tâches relatives à l'évaluation autrement longues et fastidieuses » (p. 60).

Cette question du testing adaptatif va occuper les chercheurs durant de nombreuses années et jusqu'à aujourd'hui encore. En effet, l'idée d'ajuster le niveau des questions aux besoins des élèves est un élément clé de l'apprentissage soutenu par le numérique. On notera encore que ce thème est souvent associé à la notion de test diagnostique informatisé. C'est ainsi que les travaux de Dassa (1988), Dassa et De Cotret (1994), Laurier (1996), Dechef et Laveault (1999) aborderont cette question sous différents angles. En 2003, Martin (2003) fait une bonne synthèse des recherches dans le domaine, mais, dans son bilan, il fait état d'une certaine déception liée en partie à la difficulté de gérer la complexité et les nombreux paramètres de l'apprentissage adaptatif. Grâce à la puissance des outils actuels, le testing adaptatif est de plus en plus à même de répondre aux besoins spécifiques des élèves. Un engouement important est aujourd'hui à nouveau visible autour de ce thème.

IV. L'évaluation assistée par ordinateur (EAO)

Il faut attendre le passage des années 2000 pour voir apparaître le terme d'EAO (évaluation assistée par ordinateur) dans un texte de Dionne (2006). Cet écrit est intéressant parce qu'en dépit d'un parti pris assumé pour l'EAO, il apporte pour la première fois des éléments solides et très documentés sur le mariage entre ordinateur et évaluation. L'auteur commence par expliciter clairement les liens entre les tests et l'informatique et ses plus-values. D'abord, l'EAO permet une meilleure différenciation et un meilleur respect des rythmes d'apprentissages des élèves. Ensuite, il est possible d'ajuster l'épreuve en fonction du niveau de compétences de l'élève. Dionne ajoute qu'un « des avantages manifestes de l'EAO réside dans le temps économisé lors de la correction des épreuves. (...) Un environnement informatique permet de réaliser la correction et la compilation des résultats en un temps minimal » (p. 104). Enfin, il ajoute que « dans un proche avenir, il sera possible de traiter des items à réponses élaborées grâce aux développements issus de la recherche en correction automatique de rédactions écrites » (p. 104). Même si cette possibilité n'apparaîtra que plus tard, ce texte vante aussi l'EAO pour ses capacités à fournir des feedbacks immédiats, ce qui permet aux élèves et aux enseignant·e·s d'être renseignés sur la progression de l'apprentissage.

Cela permet également de réduire le risque d'erreurs de correction et par une analyse des résultats, de cerner plus rapidement et facilement les difficultés des élèves de manière à proposer des remédiations plus adéquates. Des risques de tricheries sont évoqués par l'auteur qui note encore que l'EAO permet d'enrichir les items proposés aux élèves (recours à des situations avec des simulations pour « plonger les élèves dans une situation d'évaluation authentique » (p. 108) ou d'utiliser des laboratoires virtuels ou encore des extraits vidéo auxquels des questions sont associées). Par contre, l'auteur - fidèle aux principes fondateurs de l'ADMEE - souligne la nécessité de s'assurer de la validité des items. Cette tâche, pas si simple à réaliser, nécessite de s'astreindre tout autant que la construction de banque d'items rigoureusement élaborés en s'appuyant sur les réponses des élèves et, quand cela est possible, sur des données liées au processus de résolution de la question. L'article aborde pour terminer quelques aspects très intéressants comme les différences de validité entre les épreuves papier-crayon et électroniques, les facteurs d'appréhension liés à l'utilisation de l'ordinateur (en particulier entre filles et garçons) et des questions liées aux coûts d'implantation de ces outils (en précisant également les économies que cela pourrait amener en termes d'impression de papier).

Cette publication a été suivie de quelques autres abordant différents thèmes (Schuessler, 2009 ; Dionne, 2012 ; Béland, 2013 ; Morin, 2017 ; Drijvers, 2018 ; Sacré, 2019) qui s'égrènent jusqu'à la fin de la deuxième décennie des années 2000 où un numéro spécial de E-JIREF (*Evaluation et numérique : des pratiques éclectiques qui explorent des espaces à déchiffrer*) est consacré à cette thématique (Coen et al. 2019). Ce recueil, consécutif à un symposium organisé dans le cadre du colloque 2018 de l'ADMEE, touche de nombreux aspects comme la nature des données collectées et les modalités de récolte, les apports du numérique sur la régulation des apprentissages (diagnostic, gestion de la progression ...), la nature des feedbacks et leurs impacts sur les apprenants, la diversité des objets évalués ainsi que des aspects éthiques. Dans leur introduction, les auteurs soulignent la grande hétérogénéité des articles qui démontrent des choix très différents dans les angles d'attaque des contributeur trice s. La grande ouverture thématique de l'ouvrage se vérifie dans les papiers qui touchent autant l'engagement cognitif des étudiants lorsque l'évaluation se passe avec de très grands groupes d'étudiants (Lepage et al. 2019), les apports

technologiques dans un dispositif d'évaluation alliant des acteurs sur le terrain munis d'oreillettes et des spectateurs distants interagissant avec eux (Gardies et al. 2019), la comparaison de la quantité et la qualité des régulations entre classe réelle et classe virtuelle (Carrupt et Barras, 2019) que l'impact de l'approche par compétence, la formation à distance et l'utilisation des technologies (Leroux et al. 2019) ou encore l'approche par projet (David et Droyer, 2019) dans les dispositifs d'évaluation. À ce stade, on peut rejoindre le point de vue des auteur·e·s qui soulignent plusieurs points : d'abord le foisonnement de travaux dans ce champ et peut-être un certain manque d'axes structurants, l'amenuisement de la vague numérique observée dans le monde scolaire dès les années 2005, ensuite la résurgence de certains modèles d'évaluation « traditionnelle » renforcée par l'usage de plateformes présentant des outils relativement peu ouverts, et enfin, une ouverture vers l'utilisation plus systématique des *big data* pour permettre un retour aux « questions historiques » comme le diagnostic automatique, le profilage des apprenants, le testing adaptatif ou encore le pilotage des systèmes éducatifs par les chiffres, l'allocation des ressources fondées sur les algorithmes...

V. Evaluation formative, processus et hybridation de l'évaluation

L'évaluation formative et la régulation des apprentissages ont mobilisé un nombre très important de chercheur·e·s de la communauté ADMEE. Il n'est donc pas surprenant de voir cette thématique associée à celle des technologies. En 1990 déjà, Boulet et al. conduisent une recherche intitulée *Système informatisé pour l'évaluation formative* dans laquelle l'équipe de recherche croise des contenus musicaux avec une architecture informatique permettant à l'enseignant·e de paramétrier finement les apprentissages de ses élèves. Quelques constats sont faits, en particulier sur les avantages d'obtenir des feedbacks immédiatement et de diminuer considérablement le temps de correction en dépit d'un temps plus grand pour répondre aux questions et de problèmes d'accès aux salles informatiques.

Dans ses travaux, Campos (1990), considère « qu'un des atouts majeurs de l'enseignement assisté par ordinateur est qu'il ouvre grand les portes d'une individualisation de l'apprentissage » (p. 56). Il doit permettre d'aller plus loin qu'un simple feedback de type réponse correcte/incorrecte en proposant des remédiations spécifiques aux types d'erreurs. L'auteur évoque dans son papier différentes recherches qui s'attachent à analyser les procédures erronées des élèves, mais laisse le lecteur un peu sur sa faim quand, après de grandes démonstrations sur le nombre de procédures à prendre en compte dans l'apprentissage de la soustraction, il conclut qu'« il est probablement vain de vouloir créer un didacticiel qui tenterait d'apporter une remédiation propre à chaque type d'erreurs » parce que « de l'enseignant et l'ordinateur, le premier est le plus flexible et le plus apte à moduler les remédiations en fonction des difficultés de l'élève. Le second est moins vite écoeuré par le dépistage des erreurs et peut générer des exercices qui permettront de déceler la persistance de procédures erronées » (p. 65). Pour lui, il faut que la remédiation ait du sens et que « l'utilisation de tels outils (...) ne saurait être que profitable, dans la mesure où ils les [les enseignant·e·s] aident à se rendre compte que bien des erreurs ne sont pas dues à une simple inattention de l'élève, mais bien à une connaissance imparfaite de la matière abordée ». (p. 65). Constat qui sur le fond est déjà bien intéressant.

En matière d'évaluation formative, nous retiendrons également l'expérience originale de deux enseignants suisses (Favre et Briguet) qui, à la suite d'une conférence de Samuel Roller, développent dès 1991 un dispositif informatique pour gérer la progression de leurs élèves. Ces deux praticiens créent le logiciel Evalog (<https://evalog.ch/>) qui permet de noter très précisément les objectifs maîtrisés (ou non) par les apprenants. Les fonctions du logiciel permettent de suivre ainsi les apprentissages et, en effectuant des tris et des regroupements d'élèves selon leurs besoins, de mettre en place une différenciation pédagogique efficace. Notons qu'après de nombreuses mises à jour, le logiciel est toujours opérationnel et adapté au plan d'études suisse.

Dix ans après ces travaux, Coen (1997, 2020) publie une thèse sur l'utilisation du logiciel AutoéVal qui permet à l'enseignant et à l'élève d'accéder au processus d'écriture. L'idée de base rejoint celle précédemment évoquée d'une prise de conscience de sa propre activité

pour la réguler le mieux possible. Présentée au colloque de l'ADMEE de Mons en 1998, cette recherche explore quelques pistes intéressantes. D'abord celle d'utiliser l'ordinateur pour révéler ce qui n'est a priori pas accessible pour l'apprenant. Le logiciel produit un graphique mêlant la courbe de progression du texte à l'utilisation des outils à disposition qui sert ainsi de support pour permettre aux élèves d'expliquer leur stratégie d'écriture. Ensuite, l'analyse du protocole permet d'affiner considérablement le diagnostic des difficultés puisque l'enseignant e peut accéder aux différentes régulations effectuées par le scripteur. Et enfin, par l'ajout d'outils spécifiques et personnalisés, il permet de donner à l'élève des aides à l'écriture parfaitement adaptées à ses besoins.

Avec le développement d'Internet, le recours à des logiciels particuliers s'amenuise et l'ordinateur fait place aux plateformes numériques qui intègrent souvent des dispositifs d'évaluation ou de régulation des apprentissages. Dans cette logique, on retiendra l'article de Dionne (2012) consacrés aux systèmes à réponses personnalisées (SRP). Ces dispositifs permettent de sonder les étudiants durant les cours via un boîtier, un smartphone ou un ordinateur. Ils se présentent très souvent sous forme de quiz auxquels les étudiants doivent répondre pour donner aux enseignants les moyens d'évaluer leurs pré-acquis ou plus généralement la compréhension d'une séquence d'apprentissage. Dionne démontre que les retours donnés par les élèves sont des informations très intéressantes en particulier parce qu'elles sont immédiates et complémentaires à ce qu'il pourrait percevoir. Elles permettent de rétroagir et de proposer des ajustements de l'enseignement susceptibles de favoriser les apprentissages. Dans une revue de la littérature sur le sujet et ciblée sur l'enseignement supérieur, Detroz et Younès (2023) présentent différentes recherches récentes qui montrent l'intérêt de telles démarches pour favoriser les apprentissages profonds même si ces auteurs font état de résultats contrastés mettant en évidence l'influence de différentes variables notamment le niveau de développement pédagogique des enseignant e's concerné e's.

Cette idée est prolongée par Alvarez et al. (2021) qui proposent une plateforme numérique (www.oura2.ch) permettant de créer des micro-sondages au fil des cours ciblant - non pas les apprentissages des étudiant e's eux-mêmes - mais des variables pédagogiques liées aux

dispositifs proposés telles que l'attrait des tâches, leur valeur perçue, l'engagement des apprenants, leur sentiment de compétence, etc. L'idée centrale de cette démarche (utilisée notamment en formation) est que, dans une logique de contributions réciproques (Coen et al., 2024), les avis des élèves permettent aux professeurs d'améliorer leurs pratiques d'enseignement et de se développer professionnellement pour mieux répondre aux besoins des apprenant·e·s.

On notera enfin l'article de synthèse de Sacré et al. (2019) sur l'impact de composantes pédagogiques des plateformes numériques sur les apprentissages ou, dans un domaine plus spécifique, le travail de Loignon (2021) sur le développement d'un outil d'évaluation de la complexité linguistique.

VI. Le boum de la pandémie

Pour terminer ce petit panorama, il est évidemment impossible de passer sous silence l'impact très important de la pandémie du COVID-19. En matière d'éducation, le virage a été brutal avec l'obligation rapide d'utiliser massivement les technologies pour l'enseignement à distance lors du confinement. La communauté ADMEE n'a évidemment pas échappé à l'événement et un numéro spécial d'e-JREF a été consacré à cette période. Sans entrer dans une revue systématique des contenus de toutes ces publications, plusieurs thèmes se dégagent de ce numéro spécial et touchent de près ou de loin l'évaluation. Le premier - très représenté - aborde les nouveaux défis liés à l'évaluation à distance et les adaptations ou les transformations des dispositifs dans de nouveau contexte mobilisant de l'évaluation formative ou sommative (Burguete et al., 2020 ; Da Costa Cabral et al., 2020 ; Duroisin, 2020 ; El Hage et Nahed, 2020 ; Gilles et Charlier, 2020). Le deuxième met en évidence l'intérêt des technologies pour réduire la distance et ainsi accompagner de près la progression des apprenants en essayant de capter des traces de l'avancement de leur travail et en déployant des moyens pour les rassurer (communication asynchrone) (Coen et Pellegrini, 2020 ; El Hage et Nahed, 2020 ; Thual, 2020). Plusieurs auteurs évoquent des aspects éthiques de l'évaluation médiatisée avec les technologies en termes de stress pour les

élèves ou de relations avec les enseignants (Alonso Vilches et al., 2020 ; Da Costa Cabral et al., 2020 ; Lison, 2020). Enfin, par le biais de critères de qualité ou de questionnements ouverts, quelques textes se regroupent autour de recommandations concernant l'usage du numérique pour évaluer les apprentissages (Gilles et Charlier, 2020 ; Heilporn et Denis, 2020 ; Papi et al., 2020 ; Rached et al., 2020).

Avant d'aborder le dernier point, il nous semble nécessaire d'évoquer à ce stade l'ouvrage de Loyer et Duroisin (2024), certes publié en marge de la revue *Mesure et évaluation*, mais qui regroupe des contributions de chercheur·euse·s très actif·ve·s au sein du réseau de l'ADMEE (à commencer par les auteures du livre). Abordant successivement les questions liées aux contextes de l'évaluation, à la collecte des données, à leur analyse et leur modélisation, les coordinatrices de l'ouvrage donnent à voir un renouvellement du champ par le foisonnement de thématiques abordées.

VII. Les enjeux de l'intelligence artificielle

Depuis l'arrivée dans le grand public de ChatGPT et des autres modèles de langage, de nombreuses questions se posent en termes d'apprentissage et d'évaluation. À tel point que l'IA semble écraser toutes les autres problématiques. Le dernier colloque de 2025 (Luxembourg) a vu exploser le nombre de contributions traitant de ce thème. Alors aujourd'hui, peut-on dire que les avantages et les risques liés à l'utilisation de l'IA dans l'évaluation des apprentissages s'équilibrivent ? On peut certes soutenir que l'IA permet de gagner un temps précieux, de donner des feedbacks plus rapidement, plus personnalisés et plus adaptés, d'analyser plus finement les données liées au processus d'apprentissage et ainsi de mieux comprendre les erreurs et les difficultés des élèves. Mais on peut aussi y voir des risques importants de perte de sens des apprentissages pour les élèves, de manque d'adaptation face à des tâches moins standardisées, de manque d'empathie et de déshumanisation des interactions humaines ou encore de captation malveillante de données personnelles. Le dernier article dont nous parlerons ici aborde cinq enjeux d'évaluation face à l'émergence des IA génératives en éducation (Audran, 2024). En partant

du principe que les « IA génératives obéissent à une logique qui a l'ambition de substituer une production numérique artificielle à la création humaine », Audran pose, pour commencer, la question des compétences des utilisateurs et de leur capacité à comprendre le fonctionnement de ces outils. À la vue du nombre de paramètres mobilisés dans ces systèmes, il semble que « la seule opération évaluative que peut réaliser l'utilisateur final est d'estimer la qualité de la production à l'aune de ses propres connaissances » (p. 14). Le deuxième enjeu consiste à évaluer les illusions fondées sur la forme. En d'autres termes, il s'agit ici d'apprendre aux utilisateurs à « déceler les indices qui caractérisent les productions de ces générateurs et faire la différence entre production humaine originale et production machinique » (p. 15). Le troisième défi consiste à évaluer la part des hallucinations pour s'assurer que ce qui est produit est vrai. Cette question est cruciale et renvoie à des enjeux philosophiques (qu'est-ce que le vrai ?). Cela dit, cette aptitude reste indispensable à développer chez les utilisateurs en leur donnant les moyens de cette vérification. La dimension éthique est elle aussi très importante parce que le recours à ces outils permet à certains élèves de réaliser des tâches parfaitement conformes à ce qui est attendu. Il s'agit alors pour les enseignants d'évaluer l'authenticité des productions des élèves pour déceler d'éventuelles tricheries. Cette problématique n'est pas nouvelle, mais là, elle se pose de manière plus importante parce que sa détection est très difficile. Cette question en ouvre d'autres sur la complémentarité entre l'élève et l'IA dans la réalisation d'une tâche. Le dernier enjeu touche l'application (ou l'intégration) de l'IA pour améliorer l'évaluation. Comme nous avons pu le voir dans ce texte, l'idée de disposer d'outils capables de s'ajuster au niveau des élèves, de leur donner des tâches adaptées à leurs besoins habite les chercheurs depuis de nombreuses années. L'IA offre des perspectives intéressantes en permettant un affinage du diagnostic ou en assistant l'élève dans son autoévaluation. Dans ce sens, et à la suite d'Audran, il nous semble important de penser une intégration réfléchie, très réfléchie, de ces outils plutôt que d'organiser une résistance servile.

VIII. Constats conclusifs

Au terme de ce petit périple autour de l'évaluation et du numérique dans la communauté ADMEE, il nous semble intéressant de faire quelques constats conclusifs.

L'intérêt modéré des chercheurs pour lier évaluation et technologie ne traduit pas une volonté de mise à l'écart. L'examen des articles parus depuis plus de quarante ans nous montre au contraire une volonté d'associer avec prudence le numérique et l'évaluation. Le souci de garantir la validité des tests et de comprendre comment fonctionnent les outils semble être encore très présent dans la communauté des chercheurs de l'ADMEE. Certes, on peut voir quelques décalages entre les avancées techniques et leur utilisation pour évaluer, mais cela nous semble assez normal en regard du lien qu'entretiennent depuis toujours l'école et la société.

Nous faisons aussi le constat de l'accroissement d'une certaine diversité dans les thématiques touchant l'évaluation et la technologie. Le numéro d'E-JIREF de 2019 le montre bien. Y aurait-il un éparpillement, une dispersion excessive des intérêts, des thématiques ? Difficile de le dire. Si le testing adaptatif occupait clairement la scène dans les années 90, la diversification des outils et des usages du numérique a conduit les chercheurs à explorer des champs nouveaux, sans qu'il y ait a priori des axes forts. C'est peut-être le propre de ce domaine d'études qui par la multiplicité des usages possibles des technologies rend difficile la cohérence et l'intelligibilité.

Enfin, l'arrivée de l'IA donne l'impression d'un écrasement, comme si cette technologie s'immisçait partout jusque dans les moindres interstices : diagnostic des difficultés des élèves, remédiations et production d'exercices adaptés, jugement évaluatif, production de critères, autoévaluation et suggestion d'améliorations, analyse de données et de l'activité, profilage selon des compétences, aide à la décision... Comme le rappelle Sadin (2015), après l'âge de l'accès et celui de la mesure, nous sommes désormais dans celui de l'injonction dans lequel les machines prescrivent les comportements à suivre. Serons-nous dès lors en mesure de garder notre marge de manœuvre, de rester maîtres du jeu et de cultiver

notre esprit critique. Qu'adviendra-t-il des enfants qui n'apprendront pas comme nous à écrire, à analyser ou à créer « à l'ancienne » ? Ce regard sur le passé devrait à la fois nous rassurer, mais aussi nous ouvrir les yeux sur tous les possibles... autant les pires que les meilleurs pour toujours garantir aux élèves une évaluation intelligente et à visage humain.

IX. Bibliographie

- Alonso Vilches, V., Detroz, P., Hausman, M. et Verpoorten, D. (2020). Réception de la prescription à « basculer vers l'eLearning » en période d'urgence sanitaire - Une étude de cas. *Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 5-16.
- Alvarez, L., Çuko, K., Boéchat-Heer, S. et Coen, P.-F. (2021). Faciliter l'autorégulation de l'enseignement. Conception d'un dispositif numérique fondé sur les analytiques de l'enseignement. *Revue suisse des sciences de l'éducation*, 43(3), 366-375.
- Audran, J. (2024). Cinq enjeux d'évaluation face à l'émergence des IA génératives en éducation. *Mesure et évaluation en éducation*, 47(1), 6-26.
- Béland, S., Magis, D. et Raîche, G. (2013). Estimation des paramètres d'item et de sujet à partir du modèle de Rasch : une étude comparative des logiciels BILOG-MG, ICL et R. *Mesure et évaluation en éducation*, 36(1), 83-110.
- Boulet, M.-M., Lavoie, L., Labbé, P. et Lemay, F. (1990). Système informatisé pour l'évaluation formative. *Mesure et évaluation en éducation*, 13(1), 29-55.
- Burguete, E., Picard, N., Andrieux, N., Fourcade, L. et Perrochon, A. (2020). Évaluation par les pairs à distance lors d'un enseignement de lecture critique d'articles pour des étudiants paramédicaux. *Évaluer _ Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 41-51.

- Carrupt, R. et Barras, H. (2019). Dynamique des régulations en classe virtuelle ou en présentiel. *Évaluer - Journal international de recherche en éducation et en formation*, 5(3), 57-83.
- Coen, P.-F. (1997). Analyse des régulations d'élèves scripteurs utilisant AutoéVal, logiciel d'assistance à l'écriture. *Cahiers de la recherche en éducation*, 27(3), 261-278.
- Coen, P.-F. (2000). *A quoi pensent les enfants quand ils écrivent. Analyse des processus cognitifs et métacognitifs en jeu dans une tâche d'écriture assistée par le logiciel AutoéVal*. [thèse de doctorat en ligne]. Université de Fribourg. <http://pfcoen.ch/a-quoi-pensent>
- Coen, P.-F., Detroz, P. et Younès, N. (2019). Introduction - Evaluation et numérique : des pratiques éclectiques qui explorent des espaces à déchiffrer. *Évaluer - Journal international de recherche en éducation*, 5(3), 1-8.
- Coen, P.-F., Çuko, K. et Etienne-Tomasini, D. (2024). Évaluer et réguler les enseignements : l'utilisation de « OURA », un outil numérique pour apprécier les expériences d'apprentissage des apprenants. Dans N. Loyer et N. Duroisin (dir.), *Evaluation, apprentissage et numérique*, (p. 195-218). Peter Lang.
- Coen, P.-F. et Pellegrini, S. (2020). Une feuille de route numérique pour évaluer formativement la progression des étudiant.e.s en contexte d'enseignement à distance. *Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 59-65.
- Connes, J. (2022). *De l'IBM 360/75 au superordinateur Jean Zay*. EDP Sciences.
- Da Costa Cabral, F., Gremion, C. et Roblez, A. (2020). Évaluation formatrice, rythmes et travail de groupes à distance. *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 77-84.
- Dassa, C. (1988). L'intégration du diagnostic pédagogique aux apprentissages scolaires : de la théorie à la voie informatique. *Mesure et évaluation en éducation*, 11(1), 7-26.

- Dassa, C. et De Cotret, S. (1994). Validation d'un système informatisé de diagnostic en mathématiques au secondaire : une approche centrée sur l'analyse didactique. *Mesure et évaluation en éducation*, 16(3-4), 5-26.
- David, M. et Droyer, N. (2019). Évaluation de la co-conception d'un environnement virtuel éducatif forestier Pré-enquête à l'entrée par le critère de pertinence. *Évaluer - Journal international de recherche en éducation et en formation*, 5(3), 109-130.
- De Campos, D. (1990). Outils diagnostiques et enseignement assisté par ordinateur. *Mesure et évaluation en éducation*, 13(3), 55-67.
- Dechef, H. et Laveault, D. (1999). Le testing adaptatif par ordinateur. *Psychologie et psychométrie*, 20(2-3), 151-179.
- Detroz, P., et Younès, N. (2023). Médiations pédagogiques en jeu à l'heure des systèmes numériques interactifs : une revue de littérature narrative et critique. *Revue français de pédagogie*, 221, 119-155. <https://doi.org/10.4000/11w0f>
- Dionne, E. (2006). L'implantation de l'évaluation assistée par ordinateur (EAO) au sein du système scolaire québécois : avantages et obstacles à surmonter. *Mesure et évaluation en éducation*, 29(2), 99-116.
- Dionne, E. (2012). Les systèmes à réponses personnalisées (SRP) : un atout pour faire de l'évaluation formative en salle de classe ? *Mesure et évaluation en éducation*, 35(1), 47-65.
- Drijvers, P. (2018). Digital assessment of mathematics: Opportunities, issues and criteria. *Mesure et évaluation en éducation*, 41(1), 41-66.
- Duroisin, N. (2020). Le podcasting collaboratif, un outil pour l'évaluation formative à distance. *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 121-130.
- El Hage, F. et Nahed, R. (2020). Apprentissage par Exploitation de l'Erreur et à Distance (AEED). *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 131-142.

- Gardies, C., Faure, L. et Marcel, J.-F. (2019). L'usage d'un dispositif numérique synchrone pour l'évaluation de pratiques d'enseignement. *Évaluer - Journal international de recherche en éducation et en formation*, 5(3), 35-55.
- Gilles, J.-L. et Charlier, B. (2020). Dispositifs d'évaluation à distance à correction automatisée versus non automatisée : analyse comparative de deux formes emblématiques. *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 143-154.
- Heilporn, G. et Denis, C. (2020). Dispositifs d'évaluation en ligne : des recommandations de la littérature scientifique à des décisions précipitées en pratique. *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 165-172.
- Laurier, M. (1996). Pour un diagnostic informatisé en révision de texte. *Mesure et évaluation en éducation*, 18(3), 85-106.
- Lepage, I., Leduc, D. et Stockless, A. (2019). E-évaluation dynamique et engagement cognitif en contexte de grand groupe à l'université. *Évaluer - Journal international de recherche en éducation et en formation*, 5(3), 9-33.
- Leroux, J. L., Desrochers, M.-E. et Myre-Bourgault, M. (2019). L'évaluation des apprentissages à l'ère du numérique en enseignement supérieur : quels besoins et quels défis ? *Évaluer - Journal international de recherche en éducation et en formation*, 5(3), 85-108.
- Lison, C. (2020). La présentation orale en contexte de formation à distance : évaluer un Pecha Kucha. *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 173-180.
- Loignon, G. (2022). ALSI : un nouvel outil d'analyse automatisée de la complexité linguistique pour le français québécois. *Mesure et évaluation en éducation*, 44(3), 29-57.

- Martin, R. (2003). Le testing adaptatif par ordinateur dans la mesure en éducation : potentialités et limites. *Psychologie et psychométrie*, 24(2-3), 83-110.
- Morin, M. et Blais, J.-G. (2017). Production semi-automatisée d'une carte conceptuelle en science et technologie. *Mesure et évaluation en éducation*, 40(1), 61-99.
- Papi, C., Gérin-Lajoie, S. et Hébert, M.-H. (2020). Se rapprocher de l'évaluation à distance : dix pistes de réponse. *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 201-206.
- Rached, P., Gharib, Y. et Constantin, S. (2020). Entre le cognitif et l'émotionnel... Quelle évaluation à distance en situation de crise ? *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 223-230.
- Sacré, M., Lafontaine, D. et Toczek, M.-C. (2019). Liens entre les composantes des dispositifs d'enseignement hybride et les performances des étudiants de l'enseignement supérieur : une revue systématique. *Mesure et évaluation en éducation*, 42(3), 109-152.
- Sadin, E. (2015). *La vie algorithmique, critique de la raison numérique*. Éditions l'Echappée.
- Schuessler, K. et Forget, J. (2009). Utilisation d'une version informatisée de la didactique de précision auprès d'élèves ayant un trouble envahissant du développement. *Mesure et évaluation en éducation*, 32(1), 25-54.
- Seguin, S. P. (1984). L'utilisation des micro-ordinateurs pour l'évaluation des apprentissages : quelques perspectives. *Mesure et évaluation en éducation*, 7(1), 53-66.
- Thual, O. (2020). De l'examen écrit scientifique sur table au paramétrage d'une activité Test de Moodle. *Évaluer - Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 245-252.