

Joël Person, Christine Vidal-Gomel,
Philippe Cottier

Les Learning Analytics : trop d'analytics pas assez de learning ?

La numérisation massive des systèmes éducatifs a ouvert le champ au début des années 2010 à une nouvelle discipline : les Learning Analytics. Cet ensemble de techniques, basé sur l'exploitation des traces informatiques d'apprentissage prétend améliorer l'enseignement et l'apprentissage et les environnements dans lesquels ils se déroulent. Quels sont les apports de ces techniques pour l'éducation et ses acteurs, quelles en sont les limites ? Pour tenter de répondre à ces questions, nous avons mené une revue de littérature couvrant la période 2011-2023.

Consulter l'**Encadré méthodologique**

1. Les promesses d'une discipline émergente

1.1. Définition et origines

Les Learning Analytics (LA) recouvrent un ensemble de données, de processus et de finalités hétérogènes qui rendent leur traduction française délicate. Le mot « Learning » se traduit par apprentissage et fait ici référence à l'apprentissage humain et à l'apprentissage machine. Le mot « Analytics » désigne « *ce qui recouvre des techniques informatiques, mathématiques et statistiques pour révéler une information pertinente à partir de très larges ensembles de données* » (Djoudi et al., 2018, p. 3). La difficulté à traduire Learning Analytics en français (Djoudi, op. cit.) témoigne du fait que le champ est en voie de maturation et que le large empan de ses méthodes et objets de recherche le rend sensible à différentes interprétations et approches parfois concurrentes. Pour des raisons de commodité et parce que ce syntagme est majoritairement employé dans la littérature internationale, nous continuerons dans cette note à utiliser le terme de Learning Analytics et son abréviation : LA.

La généralisation des usages numériques dans le contexte de l'enseignement supérieur et la quantité considérable de traces informatiques qui en résulte, a contribué à l'émergence des LA définies comme « *la mesure, la collecte, l'analyse et la communication des données relatives aux apprenants et à leurs contextes dans le but de comprendre et optimiser l'apprentissage et les environnements dans lesquels il se produit* » (Siemens et Long, 2011, p.34). La nature informatique des traces produites par l'individu au cours d'une situation d'apprentissage médiatisée rend celles-ci archivables et potentiellement réutilisables à des fins d'analyse. La fouille et l'analyse des données issues du terrain permettraient ainsi d'identifier des modèles (patterns) de comportement d'apprenant en situation d'apprentissage, pouvant fournir des informations inédites aux acteurs de l'éducation. En se donnant pour objectif de suivre au plus près l'expérience d'un individu tout au long de son parcours instrumenté et en le confrontant à ses traces informatiques, ces pratiques sociotechniques ont donc une ambition transformatrice. Elles pourraient, par le truchement de systèmes de visualisation, renseigner l'apprenant (et ses enseignants) sur ses forces et points faibles dans les diverses situations d'apprentissage instrumentées rencontrées. À plus grande échelle, elles pourraient aussi fournir des informations originales sur les systèmes d'éducation et contribuer à les transformer (Buckingham-Shum, 2012).

Les LA sont nées, entre autres disciplines, de l'informatique décisionnelle, de l'Intelligence Artificielle et des statistiques (Siemens, 2013). Jusqu'à l'épidémie de COVID-19 en 2020, c'est essentiellement dans l'enseignement supérieur que sont concentrées les recherches, ce niveau conjuguant de larges cohortes d'apprenants et la mise en œuvre par les institutions de plateformes de type LMS, pourvoyeuses d'immenses quantités de données. Certaines de ces données peuvent être visualisées par diverses applications telles que des Tableaux de Bord. Préalablement incorporés dans les plateformes dédiées à l'enseignement, ces affichages fournissent des informations graphiques sur le nombre de connexions à un cours, la durée d'utilisation d'une ressource et le type d'actions effectuées, ou encore sur l'état de rendus de travaux par l'étudiant.

1.2 Le potentiel des Learning Analytics

Pour ses promoteurs, des Learning Analytics participent à une nouvelle manière d'appréhender les questions d'enseignement-apprentissage en décrivant le processus d'apprentissage instrumenté depuis de nouveaux et multiples points de vue. Dans le courant des recherches s'intéressant aux technologies pour l'éducation, l'analytique des données contribue aussi à la conception et à l'amélioration des dispositifs mobilisés en contexte d'éducation (capteurs, logiciels, plateformes). Au-delà de cette approche de recherche fondamentale documentée par les données, les chercheurs du domaine rappellent que les applications d'analytique pourraient participer à la résolution de problèmes rencontrés par les acteurs de l'éducation tels que, les taux importants d'abandon liés à la massification de l'enseignement supérieur, les inégalités sociales d'accès à l'éducation pour les élèves issus de milieux défavorisés ou encore l'orchestration d'activités complexes en classe. Le suivi des traces générées par les systèmes informatiques à un niveau individuel s'effectue par leur collecte automatisée. Les LA pourraient ainsi révéler à grain fin les pratiques en ligne de l'apprenant par des données numériques, sémantiques ou textuelles : les clics produits, les données de prêt de la bibliothèque, ses informations administratives voire ses relations interpersonnelles via les réseaux sociaux.

Les LA, lorsqu'elles sont à destination de l'apprenant, recouvrent quatre visées de traitement des données, qui opèrent selon des temporalités, des perspectives et des modalités techniques différentes. La première visée, descriptive, consiste en une restitution basique d'informations à partir

des traces disponibles de ce qui s'est passé durant la séquence documentée. La deuxième visée, diagnostique, affine la description en mettant en évidence des facteurs susceptibles d'expliquer les actions constatées. Les visées prédictives et prescriptives s'appuient quant à elles sur les actions passées de l'apprenant et la comparaison avec d'autres sujets pour envisager ses probabilités de succès ou d'échec dans le parcours d'apprentissage et, dans ce dernier cas, les propositions pour y remédier (feed-back, modules d'entraînement ou recommandation de ressources complémentaires, ...). Combinées avec les recherches en Intelligence Artificielle, les LA pourraient déboucher sur la mise en œuvre de tuteurs « intelligents » interagissant avec les apprenants en langage naturel dans une optique d'apprentissage personnalisé aussi appelé « adaptive learning ».

À une autre échelle, dans un contexte de massification de l'enseignement et de la formation en ligne, le recours aux systèmes de LA dans les plateformes d'apprentissage contribuerait également à optimiser les coûts des formations pour les organismes qui les dispensent et à épargner aux enseignants ou tuteurs des tâches répétitives.

2. Limites des LA : une dynamique de recherche freinée par de nombreux obstacles

L'enthousiasme envers ce champ en plein essor est toutefois nuancé au sein même de sa propre communauté qui, très tôt, a initié des débats très vifs pour explorer les limites et risques afférents aux LA. Ces discussions donnent lieu à une dynamique de travaux scientifiques autour de la recherche de preuves de l'efficacité des systèmes de LA, preuves difficiles à établir en raison du caractère souvent expérimental et à petite échelle des études menées, ou encore à cause d'une imprécision des concepts éducatifs mobilisés pour la conception des applications de LA.

2.1 Des limites techniques

Pour commencer, les LA sont conditionnées à l'utilisation d'appareils numériques générant des traces exploitables et accessibles, or les situations d'enseignement-apprentissage, de par leur grande diversité, n'y recourent pas toutes de manière systématique. De plus, la généralisation des usages numériques suppose des infrastructures informatiques fiables et des modalités d'accès homogènes, par exemple l'équipement personnel de l'étudiant et son environnement de travail, dont la crise du Covid a révélé qu'ils pouvaient être vecteurs d'inégalités.

Si l'on écarte cette première condition, une autre limite technique est liée à l'hétérogénéité et à l'ambiguïté des données éducatives disponibles. Une source unique de données est insuffisante si l'on considère l'apprentissage comme un processus holistique et social. Cela implique donc l'agrégation de formats de données différents, dispersés dans des services, des bases de données, détenus par des acteurs privés et publics. Cette hétérogénéité rend leur compatibilité complexe (Kitto et al., 2020) et le processus de leur traitement coûteux, en grande partie en raison du nombre important d'acteurs impliqués. Les applications de LA peuvent varier considérablement dans leurs modalités de traitement selon leur destinataire, par exemple sur des questions d'anonymisation des données, de durée de conservation ou de finalités : s'agit-il d'accompagner le parcours de l'élève, d'aider au pilotage d'un établissement, ou d'abonder des enquêtes d'évaluation internationales ?

2.2 Des limites organisationnelles

L'adoption à grande échelle de systèmes de LA représente un défi majeur lié à la complexité d'un réseau sociotechnique englobant les apprenants, les éducateurs, les administrateurs et les personnels de soutien, mais également les pratiques de ces groupes, leurs connaissances du fonctionnement de

l'enseignement-apprentissage, des technologies utilisées et de l'environnement spécifique dans lequel ils opèrent. L'introduction des LA implique donc des changements importants en termes de communication, de formation, de choix budgétaires, de suivi, etc. pour tous les groupes concernés, en fonction des spécificités locales. La généralisation de ce type de système peut être entravée par la difficulté à en établir les bénéfices pour les apprenants et les éducateurs, ces derniers en particulier doivent être associés aux projets de conception et d'implémentation, sous peine que les LA n'aient qu'un impact limité.

Alors que les institutions investissent des sommes conséquentes dans des environnements proposant ces technologies, les recherches robustes sur les effets du déploiement de solutions de LA sont rares (Motz et al. 2023). Les acteurs doivent par ailleurs être en mesure d'évaluer la pertinence d'outils d'analytique en vue de les utiliser efficacement, ce qui suppose de pouvoir s'y former. Dans certains pays se développe un enseignement fondé sur les données malgré un manque de preuves scientifiques sur son efficacité, ce qui impose une formation des enseignants à la "data-littératie", redéfinissant ainsi leur responsabilité, leur éthique et leurs relations professionnelles.

2.3 Des limites éthiques

Les algorithmes utilisés pour collecter et traiter les données informatiques ne sont pas neutres et leurs biais de conception ou la maîtrise incorrecte de leurs applications, peuvent impacter l'éducation et ses acteurs en influençant la réalité qui est mesurée. L'opacité de ces algorithmes pourrait contrarier l'acceptation de dispositifs de LA par les publics auxquels ils sont destinés, voire entraîner des plaintes et des procédures contre leurs promoteurs. Le choix d'implanter des LA est susceptible de renforcer les inégalités sociales et de désavantager certaines catégories d'acteurs par rapport aux groupes sociaux les plus représentés dans les systèmes de données (Selwyn, 2019). Les ambitions prédictives des LA, matérialisées par exemple sous forme de recommandations de ressources en fonction des profils des élèves, sont notamment mises en question quant à leur validité et aux risques qu'elles font peser sur les trajectoires des apprenants par des affichages d'indicateurs trop limités qui ne mesurent pas toujours des actions liées à l'apprentissage.

Néanmoins ces enjeux éthiques sont contrebalancés par une doctrine de l'innovation numérique portée par l'institution scolaire (MEN, 2018 ; 2023), qui voit dans les données d'éducation un potentiel en Recherche et Développement pour les acteurs de la recherche et de l'EdTech (le secteur des technologies éducatives), et ne souhaite pas l'application d'une réglementation « *trop restrictive* » (rapport IGEN-IGEAN, 2018, p. 2). Se manifeste ici une contradiction pour les décideurs, entre la nécessité éthique de protéger les données personnelles des élèves, en conformité avec le RGPD, et l'opportunité économique que représente leur exploitation. Dans la course à l'innovation qui se joue entre les états, les chercheurs en informatique doivent pouvoir disposer de « matière première » – les données – afin d'entraîner leurs algorithmes et, peut-être, donner naissance à de nouvelles « licornes ». Ce contexte de compétition internationale fait du dépôt de brevets industriels issus de la recherche un enjeu stratégique pour le monde académique, enjeu qui irrigue les recherches sur les LA.

2.4 Des limites épistémologiques

L'acception de l'apprentissage contenue dans le terme « Learning » n'est pas évidente puisqu'elle concerne ici, comme nous l'avons déjà mentionné, l'apprentissage humain et l'apprentissage machine. Si l'on reste du côté de l'humain, il n'existe pas de définition univoque de l'apprentissage. Les traces, en documentant des situations particulières, spécifiques à des catégories d'apprenants et à leur

contexte, tendent à segmenter et réduire la notion même d'apprentissage à une suite de micro-tâches. Que mesurent les LA et que peuvent-elles dire du processus d'apprentissage ? Des chercheurs pointent le risque de déséquilibre entre « analytics » et « learning » où l'approche computationnelle, à visée essentiellement comptable, l'emporterait sur une approche émancipatrice centrée sur l'apprenant et son développement (Motz et al., 2023). Cette approche plus quantitative questionne par conséquent les contenus mêmes des situations d'apprentissage, ceux-ci (scénarios, parcours, ressources, modes d'évaluation) ne risquent-ils pas d'être modélisés par les ingénieurs pédagogiques en fonction du type de traces disponibles sur les plateformes d'enseignement ? Or, une partie des outils de LA actuellement déployés repose sur l'architecture de systèmes de formation en ligne, dont la conception est trop rarement soutenue par les apports des Sciences de l'éducation et des théories de l'apprentissage (Viberg et al., 2018).

L'exploitation et l'analyse des données liées aux situations d'apprentissage supposent une activité individuelle et collective de l'apprenant essentiellement médiée par les outils numériques, or il s'agit là d'une vision réductrice de ce qu'est le processus d'apprentissage. Cette vision ne peut restituer la richesse des activités (cognitives, sociales, biologiques, émotionnelles) d'un individu, des interactions et médiations dont il peut bénéficier, et des éléments de contexte liés à son expérience.

Conclusion

À l'issue de cette revue de littérature, il apparaît que les Learning Analytics, par leur ancrage principal dans les sciences informatiques, s'appuient davantage sur les opportunités d'innovation promises par l'abondance de données produites en contexte éducatif, plutôt que sur les résultats de la recherche en éducation et l'intérêt des professionnels de l'éducation relativement à l'exercice de leur métier. Autrement dit, le développement de ces techniques relève essentiellement d'une approche technocentrée (Rabardel, 1995). Les limites de la recherche informatique sur le thème des traces des apprenants résident principalement dans un manque de connaissance de l'activité des acteurs, en particulier pour ce qui touche à l'apprentissage humain, dont la modélisation reste problématique (Linard, 2001) et souvent ancrée dans un modèle Béhavioriste (Legros et al., 2002). Si nous considérons l'activité des individus comme principale unité d'analyse de leur apprentissage, le traçage des seuls artefacts numériques, en raison de ses caractéristiques (limité, localisé, fragmenté), ne peut suffire à documenter les parcours d'apprentissage et/ou d'enseignement. Les données produites de cette façon n'ont de sens que couplées au regard « entraîné » de l'enseignant, à celui de l'apprenant ou de tout autre acteur éducatif, capable de les expliciter à partir d'informations contextuelles (Person, 2024).

Ainsi, la création d'artefacts de type Tableau de bord à partir de cet objet « flou » que sont les Learning Analytics nécessite, comme le préconise la littérature, une participation des professionnels de l'éducation à la démarche de conception des outils qui les concernent. Cette participation peut contribuer, dès l'amont du projet de conception, à éclaircir la notion de traces d'apprentissage, ce qu'elle signifie pour les différentes parties prenantes, ainsi que les éventuels bénéfices que pourraient en retirer les enseignants et leurs élèves en regard de leurs visées respectives. Encore faut-il que la démarche participative soit réellement "centrée utilisateur" : qu'elle se préoccupe de l'activité des acteurs et pas seulement de ce qu'ils peuvent en dire hors contexte (Person, 2024).

Les LA, par leur irruption dans le champ de la recherche en éducation et les questions qu'elles soulèvent pour les acteurs sur les plans pédagogiques, didactiques, juridiques, et déontologiques, représentent donc un nouvel objet qu'il convient de questionner, en faisant un pas de côté pour éviter le « techno-solutionnisme » ambiant. Quels sont les acteurs de l'éducation qui sont alors efficacement

outillés ? Quels sont leurs objectifs ? Et quels en seraient les impacts pour d'autres acteurs du systèmes, enseignants et apprenants notamment ?

Joël Person

Docteur en sciences de l'éducation et de la formation, CREN, Thème 1, Nantes Université

Christine Vidal-Gomel

Professeure des universités, CREN, Thème 1, Nantes Université

Philippe Cottier

Maître de conférences, CREN, Thème 1, Nantes Université

Encadré méthodologique

Contrairement à une revue systématique qui se construit à partir de questions spécifiques établies en amont et d'une recherche d'exhaustivité sur le sujet investigué, notre approche a été celle d'une revue de littérature « narrative » (Baumeister et Leary, 1997). La revue de littérature narrative ne s'appuie pas sur une question de départ précise. Elle vise davantage à obtenir une vue d'ensemble de ce que dit la recherche scientifique dans les ouvrages et articles publiés sur un sujet ou un thème précis. Elle permet ainsi de rassembler un vaste corpus et de relier entre elles différentes études en lien avec le thème exploré sans les contraintes de sélection d'une revue systématique. Nos questions plutôt générales portaient sur la définition des Learning Analytics, les processus de traitement des données, leurs usages pour l'éducation.

Notre travail de « cartographie » s'est constitué à partir d'articles, de dossiers thématiques, d'actes de colloques et de conférences, et identifie au cours de la décade écoulée (2011/2023) les questions vives sur les dimensions méthodologiques, théoriques et parfois politiques de l'analytique des données d'apprentissage. Plus de 80 références internationales ont finalement été mobilisées dans la thèse.

Nous nous sommes efforcés de réduire les biais de sélection en incluant les méta-analyses ou les revues de littératures systématiques en lien avec les LA et les Tableaux de bord. Nous avons également privilégié un suivi de chercheurs particulièrement actifs et reconnus dans le domaine, notamment par les références à leurs publications relayées dans la communauté. Les travaux du groupe GtNum 2 de la DNE « Analyse des traces d'apprentissage - Learning Analytics » et le dossier que Daniel Peraya a consacré à ce thème en 2019 pour la revue Distances et Médiations des Savoirs ont permis d'en identifier plusieurs. Pour l'international, on peut notamment citer Dragan Gašević ou Simon Buckingham-Shum, auteurs prolifiques, présents dès l'origine des travaux sur les LA, toujours engagés dans les avancées actuelles du champ. Leurs publications sont fréquemment citées dans la littérature. En France, des chercheuses telles que Vanda Luengo, Margarida Romero ou Anne Boyer mènent elles aussi des projets sur les données d'apprentissage et sont des chevilles ouvrières de la communauté EIAH (Environnement Informatique pour l'Apprentissage Humain).

Bibliographie

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of general psychology*, 1(3), 311-320.

Buckingham Shum, S. (2012). Learning analytics policy brief. UNESCO Institute for Information Technology in Education.

Djoudi, M., Luengo, V., El Kechai, H., Cerisier, J.-F., Maugard, E., Cherigny, F., ... Beust, P. (2018). *Thème 2 : Learning Analytics. Axe 6 : Terminologie du Learning Analytics*. Paris : MENESR.

IGEN-IGAENR (2018). *Données numériques à caractère personnel au sein de l'éducation nationale*. Rapport 2018-016.

Kitto, K., Whitmer, J., Silvers, A., & Webb, M. (2020). Creating data for learning analytics ecosystems. *SoLar Position Papers*. <https://www.solaresearch.org/publications/position-papers/>.

Legros, D., Maître de Pembroke, E., & Talbi, A. (2002). Les théories de l'apprentissage et les systèmes multimédias. Dans Denis LEGROS, Jacques CRINON (dir.), *Psychologie des apprentissages et multimédia* (pp. 23-39). Paris : Armand Colin.

Linard, M. (2001). Concevoir des environnements pour apprendre : l'activité humaine, cadre organisateur de l'interactivité technique. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 8(3), 211-238.

Motz, B., Bergner, Y., Brooks, C., Gladden, A., Gray, G., Lang, C., ... & Quick, J. (2023). A LAK of direction: Misalignment between the goals of learning analytics and its research scholarship. *Journal of Learning Analytics*. DOI : 10.18608/jla.2023.7913

Person, J. (2024). *Conception participative de learning analytics croisant études d'usages et analyse de l'activité. Le cas d'un projet réunissant enseignants de lycée et chercheurs*. Thèse de doctorat. Nantes Université.

Rabardel, P. (1995). *Les hommes et les technologies ; approche cognitive des instruments contemporains*. Armand colin.

Selwyn, N. (2019). What's the Problem with Learning Analytics?. *Journal of Learning Analytics*, 6(3), 11-19.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in human behavior*. <https://www.sciencedirect.com/science/article/pii/S0747563218303492>