



**HAL**  
open science

## L'analytique des apprentissages numériques

Hugues Labarthe, Vanda Luengo

► **To cite this version:**

Hugues Labarthe, Vanda Luengo. L'analytique des apprentissages numériques. [Rapport de recherche]  
LIP6 - Laboratoire d'Informatique de Paris 6. 2016. hal-01714229

**HAL Id: hal-01714229**

**<https://hal.science/hal-01714229>**

Submitted on 21 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# L'ANALYTIQUE DES APPRENTISSAGES NUMERIQUES

Hugues LABARTHE, Vanda LUENGO

Décembre 2016

## Contributeurs

Hugues LABARTHE.

Docteur en histoire médiévale, Hugues Labarthe est professeur agrégé d'Histoire au Lycée Delacroix, à Drancy, dans l'académie de Créteil. Durant son master en *Ingénierie et Management de la Formation en Ligne*, obtenu en 2015 à l'Université Pierre et Marie Curie, il a conçu une application pour l'apprentissage du croquis au baccalauréat général (<http://Croquiz.com>) ainsi qu'un espace pédagogique pour l'accompagnement des Lycéens en histoire-géographie (<http://Tissmi.com>). Investi dans la réalisation d'un système de recommandation pour améliorer la persistance des apprenants sur une plateforme de MOOC, il a effectué deux séjours de formation en fouille de données éducatives au CHAI (*Computer Human Adapted Interaction Research Group, University of Sydney*) en 2015 et 2016. Il a co-publié les résultats de ces recherches dans les actes des conférences EMOOCS 2016 (Graz, Autriche) et EDM16 (Raleigh, NC).

Vanda LUENGO

Vanda Luengo est Professeure à Sorbonne Universités, UPMC Paris 6. Elle occupe la chaire Environnements Informatiques pour l'Apprentissage Humain (EIAH) au Laboratoire d'Informatique de Paris 6 (LIP6) et dirige l'équipe MOCAH. Ses recherches portent sur le diagnostic des connaissances, la modélisation du feedback épistémique en EIAH et sur les *learning analytics*. Responsable scientifique du projet ANR Hubble (*Human oBservatory Based on analysis of e-learning traces*), elle coordonne des équipes de recherches pluridisciplinaires dans le but de développer un observatoire national pour la construction et le partage de processus d'analyse de données massives, issues des traces laissées dans des environnements de type *e-Learning*.

Ce rapport a été rédigé à l'initiative de la Direction du Numérique pour l'Education (DNE-MEN).

Nous remercions Nicolas Balacheff pour sa relecture attentive.

## Sommaire

Introduction .....	4
1 L'émergence de communautés scientifiques internationales .....	5
1.1 Analytics, Data Sciences & Data mining .....	6
1.2 Données d'apprentissage : deux jeunes communautés scientifiques en cours d'hybridation .....	8
1.3 Spécificités et objectifs des Learning Analytics .....	11
2 Questions de recherche en science des données de l'apprentissage .....	13
2.1 Prédire la progression de l'apprenant .....	14
2.2 Mesurer les interactions sociales.....	16
2.3 Analyser le discours .....	18
2.4 Donner à voir l'apprentissage.....	20
3 Les enjeux éthiques et déontologiques des <i>Learning Analytics</i> .....	22
3.1 Sur le marché éducatif, le mouvement open-source.....	23
3.2 L'expérimentation: entre éthique, droit et évolution.....	25
3.3 Les Learning Analytics en 2025 : huit scénarios d'évolution.....	28
Conclusion. L'Analytique des apprentissages numériques : une équation à trois inconnues.....	31
Annexes.....	32
Bibliographie .....	34

## Glossaire

Glossaire 1. Définitions usuelles en Science des Données.....	6
---	---

## Tables

Tableau 1. EDM et SoLAR : deux approches complémentaires.....	9
Tableau 2. Niveaux, objets d'analyse et acteurs en Academic & en Learning Analytics d'après LONG, SIEMENS 2011, p. 4.....	11
Tableau 3. Des objectifs communs à EDM et aux Learning Analytics, d'après BAKER 2014, p. 257-262	11
Tableau 4. Objectifs, techniques et apports des modèles de prédiction .....	14
Tableau 5. L'actualité de la recherche en <i>Social Learning Analytics</i> .....	16
Tableau 6. L'actualité de la recherche en Analyse de discours .....	18
Tableau 7. Evaluer l'utilité des Tableaux de bord pour l'apprentissage.....	20
Tableau 8. Les Learning Analytics (LA) en 2025 : huit scénarios à l'étude .....	30
Tableau 9. Présidence scientifique des conférences EDM ( <i>Program Chairs</i> ).....	32
Tableau 10. Présidence scientifique des Conférences LAK.....	32
Tableau 11. Chercheurs par pays dans les <i>Program Committees</i> d'EDM et LAK (2009-2016) .....	33
Tableau 12. Top 10 des universités européennes en nombre de <i>reviewers</i> aux conférences EDM et LAK33	
Tableau 13. Auteurs publiés dans les deux revues JEDM & JLA, de 2009 à 2016 .....	33

## Figures

Figure 1. Les <i>reviewers</i> des <i>Program Committees</i> des conférences EDM et LAK .....	10
Figure 2. Graphes d'analyse de l'activité collaborative .....	17
Figure 3. Du marquage à l'analyse statistique de l'activité collaborative sur <i>Cohere</i> .....	19
Figure 4. LARAe, un tableau de bord pour maîtriser les données d'apprentissage .....	21
Figure 5. Faisable et/ou désirable ? L'évaluation des scénarios .....	29

# Introduction

Evaluer les capacités d'abstraction des apprenants, détecter leur perte d'attention, adopter une pédagogie différenciée, dresser un bilan personnalisé actualisé au fil de l'apprentissage : voici autant de tâches qui reposent sur la capacité d'un enseignant à observer, analyser et réinvestir les traces comportementales et cognitives d'un apprentissage. Bien malgré lui, ce professionnel ne capte qu'une partie infime de ces données ce qui limite ses possibilités d'interprétation d'un geste, d'un exercice inabouti, d'une erreur de réappropriation. Comme dans d'autres domaines, l'observation humaine non instrumentée est limitée et fragile.

Avec le glissement des activités d'apprentissage vers des dispositifs numériques, ces traces changent de statut : en temps réel ou en différé, à distance ou en présentiel, elles n'ont jamais informé, de façon aussi fine et massive, l'écart entre le dire et le faire. Tableaux numériques, ordinateurs, tablettes, liseuses, smartphones sont susceptibles de capter toujours plus de données sur ce qui est *en train* de se jouer, sur le plan verbal et comportemental, dans un processus d'apprentissage. Produire, collecter, analyser et réinvestir ces traces numériques permettrait d'aider les acteurs de la communauté éducative – apprenants, parents, personnels d'éducation, enseignants, gestionnaires et administrateurs – dans les enjeux auxquels ils doivent, chacun, faire face dans la perspective du socle commun de connaissances, de compétences et de culture, qui requiert désormais d'évaluer à parts égales la maîtrise de la langue, les connaissances disciplinaires et les capacités d'autonomie et d'initiative.

Révéler ce qui se joue dans un processus d'apprentissage est l'enjeu des *Learning Analytics* (LA). A la croisée des Sciences Sociales et de l'Informatique, une communauté scientifique s'emploie depuis une dizaine d'années à développer des technologies et des techniques pour mieux comprendre les ressorts de l'apprentissage, de façon à en améliorer l'accompagnement et l'environnement notamment par des dispositifs informatiques adaptables et adaptés. Quelle est l'ampleur de ce mouvement ? En quoi l'introduction de nouveaux modèles, méthodes, visualisations, algorithmes peut-elle renouveler notre conception de l'apprentissage ? Quels sont enfin les grands débats qui accompagnent l'introduction de ces dispositifs dans l'enseignement ?

# 1 L'émergence de communautés scientifiques internationales

De la salle de cotation au supermarché, du laboratoire de physique quantique aux recherches sur internet d'un utilisateur ordinaire, notre société semble désormais régie par l'informatique, la recherche opérationnelle et la statistique. *Big data*, intelligence artificielle, *data mining*, *Analytics* sont ainsi devenus des termes usuels pour penser et animer notre monde. Le domaine éducatif est longtemps resté étranger à cette mutation. Cependant de nombreuses communautés œuvrent dès les années 1990 à la fouille de données éducatives, la recherche sur les environnements informatiques pour l'apprentissage humain, la collaboration assistée par ordinateur.

En 2007, la communauté scientifique IEDMS (pour *International Educational Data Mining Society*) émerge, puis une autre avec la création de SoLAR (*Society for Learning Analytics Research*) en 2011. Au cœur de la recherche sur les *Learning Analytics*, ces deux communautés développent des méthodes, des approches et objectifs sensiblement différents. Pourtant, au fil de leurs conférences internationales et revues spécialisées respectives, leurs états-majors ont œuvré pour un rapprochement toujours plus étroit de leurs communautés et travaux.

Au terme de cette première partie, nous rappelons ce qui fait la spécificité d'une analytique de l'apprentissage vis-à-vis de données éducatives mieux connues sous la dénomination *Academic Analytics*. Nous présentons enfin les objectifs, méthodes et travaux de référence de ce domaine de recherche dont l'expertise des traces vise à expliquer l'apprentissage.

## 1.1 Analytics, Data Sciences & Data mining

Le terme *Analytics* désigne de façon usuelle des techniques informatiques, mathématiques et statistiques pour révéler une information pertinente à partir de très larges ensembles de données. Par extension, les *Analytics* permettent sur la base d'actions réalisées, de comprendre, voire de prédire, le potentiel de futures actions dans une quête de performances et d'efficacité. Si de nombreux praticiens francophones utilisent le néologisme *Analytique* pour nommer leur discipline, il reste cependant inconnu des dictionnaires généralistes.

Les méthodes employées sont issues du champ de la Science des Données (*Data Science*), dont le syntagme apparaît en 1996 sous la plume de Chikio Hayashi [HAYASHI 1998]. En 2003 est lancé le *Journal of Data Science*. Ces méthodes se sont développées à compter des années 60 à la croisée des statistiques et de l'informatique, comme en témoignent les travaux de John W. Tukey ou de Peter Naur [TUKEY 1977, NAUR 1969]. Popularisées en 1994 par le succès du *data mining* en *marketing*, elles ont plus profondément innervé le champ scientifique : ainsi, la revue *Computers in Biology and Medicine* devient-elle pionnière en *Analytics* dès sa création en 1970 [BAKER, SIEMENS 2014]. Certains chercheurs prétendent rassembler les sciences formelles sous l'étendard de cette nouvelle science des données [ZHU 2015]. De fait, le principe même du *data mining* est étranger aux théories sur le comportement humain comme la linguistique, la psychologie ou la sociologie : la structure des données serait consubstantielle aux données elles-mêmes. On ne peut pourtant pas réduire les *Analytics* à une opposition entre statistique probabiliste et algorithmes de modélisation des données *ab nihilo*. L'enjeu des *Analytics* est de dépasser la simple description et extrapolation, pour leur substituer la modélisation, la recommandation et la prédiction [DAVENPORT *et al.* 2010]. En visant une prise de décision motivée par les données elles-mêmes (*data driven decision-making*), les *Analytics* sont considérées améliorer la productivité et les résultats des organisations [BRYNJOLFSSON *et al.* 2011].

### Glossaire 1. Définitions usuelles en Science des Données

- *Big data* : désigne un jeu de données massives, dont la taille excède celle de capture, de conservation, de gestion et d'analyse des outils logiciels de bases de données typiques. Par contraste, on désigne par l'expression *thick data* des données qualitatives recueillies et analysées pour trouver le sens d'un phénomène.
- *Data mining* : en recherche fondamentale, application d'algorithmes, issus de la statistique ou de l'intelligence artificielle, pour l'extraction de l'information (utile et inconnue) de gros volumes de données non structurées.
- *Business intelligence* : méthodes d'analyse et de modélisation pour anticiper des *scenarii* liés au fonctionnement d'une organisation.
- *Knowledge Discovery in Databases (KDD)* : le processus de préparation, de sélection, d'apprêt et finalement d'interprétation des données par les techniques du *data mining* [FAYYAD *et al.* 1996]. L'apparition d'une nouvelle revue *Data Mining and Knowledge Discovery* en 1997 consacre l'ascendance de l'expression *data mining* sur KDD.

Si le syntagme *Learning Analytics* ne s'impose qu'en 2011, les données d'apprentissage sont déjà très intensivement mobilisées par les équipes pluridisciplinaires œuvrant pour la conception d'environnements informatiques pour l'apprentissage humain (EIAH), parmi lesquels les tuteurs intelligents. La modélisation a priori de l'apprenant, l'analyse exploratoire de données [BENZECRI 1973], la création du groupe *Intelligence artificielle et Didactique* [BALACHEFF 1994], puis les conférences internationales AIED (*Artificial Intelligence in Education*) et francophones EIAO/EIAH ont posé les fondements de problématiques et de méthodes réinvesties par les *Learning Analytics*. Quatre facteurs expliquent plus largement leur développement [BAKER, SIEMENS 2014] :

- Le volume des données mis à disposition des chercheurs a très rapidement progressé : soit grâce à des archives publiques comme le *Pittsburgh Science of Learning Center DataShop* (<https://pslcdatashop.web.cmu.edu/>) ; soit par la multiplication des dispositifs d'apprentissage en ligne (LMS pour *Learning Management System*, MOOCs pour *Massive Open Online Courses*) permettant de capter les interactions des utilisateurs ; soit par croisement avec des données académiques.
- Ces données sont davantage structurées et utilisables : les travaux se succèdent en terme d'interopérabilité des données venant de plusieurs environnements d'apprentissage [WALKER 2012, NIEMANN *et al.* 2013].
- Les capacités de calcul des smartphones d'aujourd'hui dépassent celles des ordinateurs d'il y a 10 ans.
- Enfin de nouveaux frameworks comme Apache Hadoop permettent de gérer des données à la mesure du web ; et de nombreux outils d'analyse, adaptés de la *Business intelligence* à l'éducation, permettent de mener des recherches sans être nécessairement avancé en programmation ou en sciences statistiques. Rapid Miner (<https://rapidminer.com/>), R (<https://www.r-project.org/>), KEEL (<http://www.keel.es/>), Weka (<http://www.cs.waikato.ac.nz/ml/weka/>), SNAPP (<http://www.snapsurveys.com>) incluent ce type d'algorithmes.

De quelles données parle-t-on ? De données qualitatives et de traces : les données qualitatives correspondent par exemple aux réponses à des formulaires des usagers alors que les traces regroupent l'ensemble des interactions d'un usager avec son environnement d'apprentissage. Sur les principales plateformes de *e-Learning* - LMS et plateformes de MOOCs -, les analystes fouillent donc données et traces de la présence d'un apprenant (logs sur les pages, durée de consultation) ; la complétion du programme de formation (lecture des pages de cours, tentatives aux quizz) ; le succès (scores obtenus) ; la participation (posts et messages). Ces données sont principalement collectées selon la norme SCORM (*Shareable Content Object Reference Model*) initialement implémentée pour assurer l'interopérabilité des documents. D'aucuns pourraient juger que la granularité des données désormais engrangées est bien moins fine que celles mobilisées dans les expérimentations antérieures sur les tuteurs intelligents lorsqu'elles permettaient d'évaluer la concentration, les émotions et les capacités cognitives d'un apprenant via des capteurs physiologiques, l'analyse automatique des expressions faciales ou des instruments d'auto-évaluation [CALVO, D'MELLO 2011; D'MELLO, GRAESSER 2012]. En l'espace de 2-3 années, environnements, problématiques et méthodes ont changé : on est passé de données *a priori* destinées à évaluer certains aspects émotionnels, comportementaux ou cognitifs, à des données tout venant mais massives à la fois par le nombre des utilisateurs, l'intensité des usages d'un utilisateur particulier et l'entrée du monde réel dans le monde numérique *via* la géolocalisation et les indicateurs biométriques. La masse croît exponentiellement et les techniques permettent de structurer *a posteriori* ce qui est pertinent.



## 1.2 Données d'apprentissage : deux jeunes communautés scientifiques en cours d'hybridation

La première occurrence du syntagme *Learning Analytics* sur la toile provient d'une étude de marché australienne pour la mise en place d'une formation professionnelle en ligne, en 2000 [MITCHELL, COSTELLO 2000]. Les initiatives sont pourtant bien antérieures et multiples. La discipline émerge sous le nom d'*Educational Data Mining* à la conférence ITS (*Intelligent Tutor Systems*) de Montréal en 2000 [GAUTHIER *et al.* 2000] ; en 2005, un premier atelier intitulé *Educational Data Mining* est organisé à Pittsburgh conjointement à la conférence AAAI (*Association for the Advancement of Artificial Intelligence*). Ces travaux sont alors conduits sur des données issues de systèmes de tuteurs intelligents [KOEDINGER, CORBETT 2006]. Ensuite, Romero et Ventura publient un état de l'art en fouille de données éducatives [ROMERO, VENTURA 2007]. Depuis 2008 EDM désigne une conférence internationale annuelle. Vient alors la première revue consacrée à la fouille de données dans les sciences de l'éducation, le *Journal of Educational Data Mining* en 2009. En 2011, la *Society for Learning Analytics Research* crée la revue *Journal of Analytics*. Les rencontres se succèdent [BAKER, SIEMENS 2014 ; cf. Tableau 9 et 10]. Mais ces deux communautés cultivent leurs spécificités [BALACHEF, LUND 2013]. Elles s'emploient cependant à fédérer leur propre domaine par des publications généralistes [ROMERO *et al.* 2011, SAWYER 2014a, GASEVIC *et al.* 2015]. Elles trouvent des alliées dans les communautés suivantes : *Artificial Intelligence in Education (AIED)*, *Intelligent Tutoring Systems (ITS)*, *User Modeling and Adaptive Personalisation (UMAP)*.

A EDM comme à SoLAR, les chercheurs partagent un même enjeu : améliorer les expériences d'apprentissage et leur environnement. Pourtant, ils définissent différemment leur champ d'expertise. EDM se définit en 2009 comme « un domaine de recherche impliquant : 1. Le développement de méthodes dédiées à l'exploration des spécificités des données d'apprentissage ; 2. Le déploiement de ces méthodes pour une meilleure compréhension des apprenants et de l'environnement d'apprentissage » [BAKER, YACEF 2009, p. 4]. En réponse, Georges Siemens définit en 2011 les *Learning Analytics* comme « l'évaluation, l'analyse, la collecte et la communication des données relatives aux apprenants, leur contexte d'apprentissage, dans la perspective d'une compréhension et d'une optimisation de l'apprentissage et de son environnement » [LONG, SIEMENS 2011].

Quelles sont ces différences ? Spécialisés en fouille de données, la communauté EDM priorise le développement de nouveaux outils et algorithmes pour révéler les modèles enfouis dans les données d'apprentissage. Ces modèles touchent à des micro-concepts impliqués, par exemple, dans l'apprentissage du calcul. Dans la communauté EDM, la conception d'algorithmes vise ainsi à donner au logiciel la capacité de prédire les résultats d'un apprenant et de personnaliser sa stratégie d'apprentissage. Dans le sillon de SoLAR en revanche, modélisation et visualisation des données sont transmises aux acteurs de l'apprentissage (apprenant, personnels d'éducation, enseignants, etc.). C'est là la différence fondamentale entre EDM, dont le produit de la recherche alimente une machine, et SoLAR qui vise à amplifier le rôle décisionnel des acteurs de l'apprentissage. Les *Learning Analytics* ne sont pas une fin *per se*, mais deviennent un moyen, pour la communauté éducative, d'améliorer l'expérience d'apprentissage. Le véritable enjeu est alors de ne pas noyer les acteurs dans les données mais de leur donner des moyens pertinents de les explorer [CHARLEER *et al.* 2014a].

**Tableau 1. EDM et SoLAR : deux approches complémentaires**

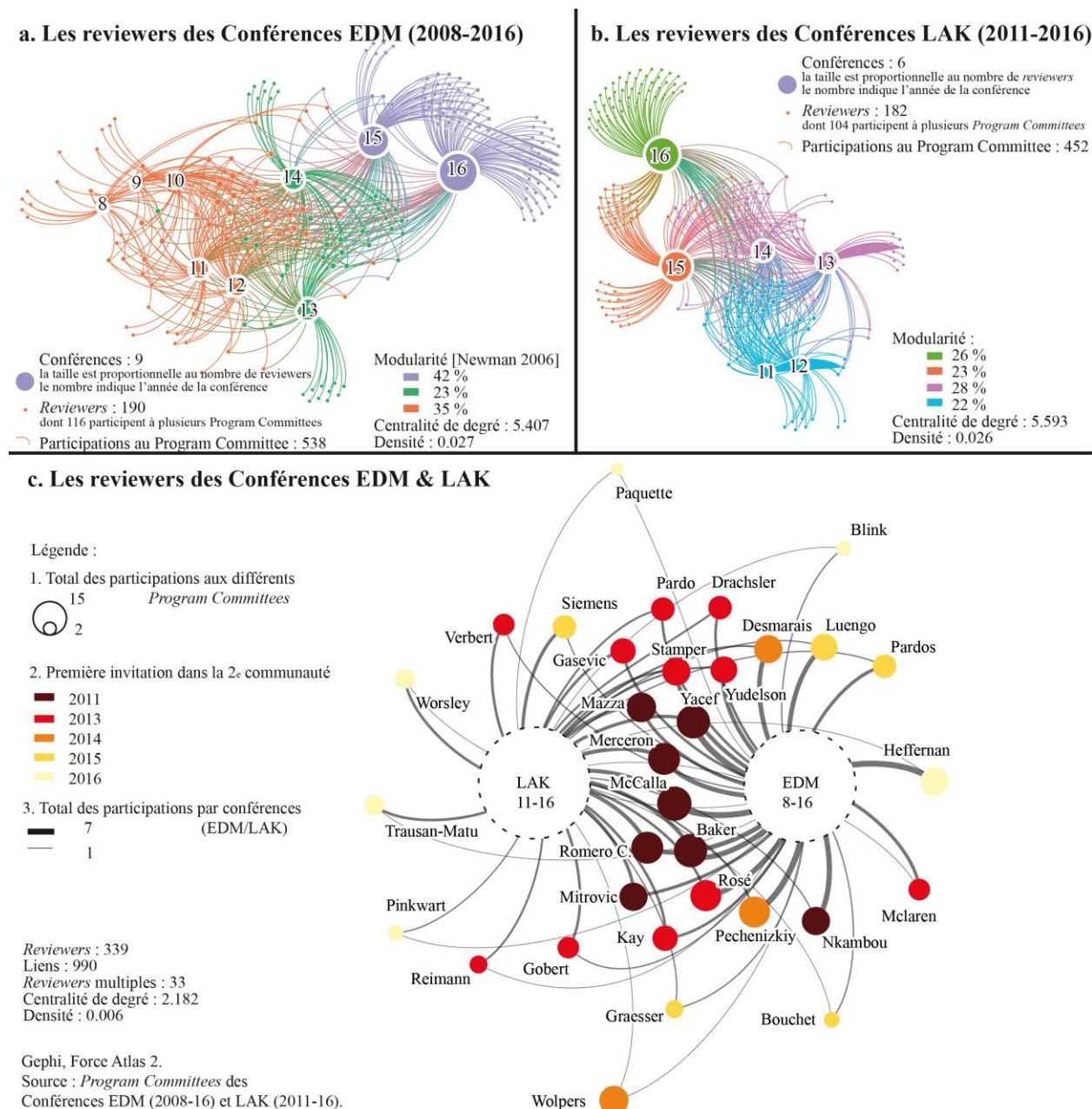
Acronyme	EDM	SoLAR
Intitulé	<i>International Educational Data Mining Society</i>	<i>Society for Learning Analytics Research</i>
URL	<a href="http://www.educationaldatamining.org/">http://www.educationaldatamining.org/</a>	<a href="http://www.solaresearch.org/">http://www.solaresearch.org/</a>
Création	2007	2011
Définition du domaine des Sciences des données éducatives	<i>“discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in”</i> . [BAKER, YACEF 2009, p. 4]	<i>“the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs”</i> [LONG, SIEMENS 2011]
Méthodes privilégiées	Fouille de données automatisée : méthode de découverte automatisée en vue d'établir les meilleures prédictions possibles.	Outils d'aide à la décision : méthode de découverte des données par les acteurs de l'apprentissage, via des modules d' <i>Analytics</i> et de visualisation.
Approche	La recherche en fouille de données éducatives a de très forts appuis théoriques et procède en réduisant le système d'apprentissage à ses composantes principales, en modélisant séparément les apprenants, les tuteurs, le domaine enseigné etc.	La recherche en <i>Learning Analytics and Knowledge</i> privilégie une approche systémique, appréhendant la situation d'apprentissage comme un ensemble. Elle ne repose pas sur des modèles théoriques en Sciences Humaines et Sociales.
Objectif de recherche	Concevoir des systèmes automatiques avec des objectifs de suivi, d'adaptation, de personnalisation Cf. ARROYO <i>et al.</i> 2007, BAKER <i>et al.</i> 2006, CORBETT & ANDERSON 1995.	Donner de l'autonomie aux acteurs de l'apprentissage (enseignants et élèves). Cf. ARNOLD 2010.
Revue	<i>Journal of Educational Data Mining</i> , 2009- <a href="http://www.educationaldatamining.org/JEDM/">http://www.educationaldatamining.org/JEDM/</a>	<i>Journal of Learning Analytics</i> , 2014- URL : <a href="http://learning-analytics.info/">http://learning-analytics.info/</a>
Conférences	<i>International Educational Data Mining Conferences</i>	<i>Conference on Learning Analytics and Knowledge (LAK)</i>

Quelle place joue la recherche européenne dans ces communautés ? L'animation de ces conférences revient principalement à la communauté anglo-saxonne. La présidence scientifique des conférences EDM et LAK revient pour moitié à des chercheurs états-unis (EDM : 9 chaires sur 18) ou nord-américains (LAK : 7 chaires sur 14) [Tableaux 9 et 10, OCHOA *et al.* 2014]. Plus largement, sur 339 chercheurs chargés d'évaluer les articles soumis à chacune de ces 15 conférences (*reviewers*), 45% appartiennent à une université nord-américaine, 41% exercent en Europe. La proportion des *reviewers* européens est plus importante dans les conférences LAK (42% contre 35% aux conférences EDM). Parmi les 10<sup>ères</sup> universités européennes figurent celles du Mans et de Grenoble et plus récemment Paris 6 du fait de la mobilité des chercheurs [Tableau 12].

Peut-on distinguer au sein des communautés EDM et LAK une évolution dans le pilotage des conférences ? Un taux similaire de *reviewers* intervient d'une année à l'autre : 61% (soit 116 *reviewers*) interviennent plus d'une fois aux *Program Committees* EDM, 57% pour LAK. Le réseau d'affiliation des *reviewers* EDM de 2008 à 2016 souligne ce fait (Figure 1a) : 190 nœuds correspondant aux *reviewers* sont reliés par 538 liens aux 9 conférences (notées de 8 à 16 en fonction du millésime). Sur les *Program Committees* LAK, les nombres sont semblables (Figure 1b) : sur 182 *reviewers* et 452 liens aux six conférences notées de 11 à 16, 104 *reviewers* interviennent de 2 à 6 fois. Mesurer la modularité de ces deux graphes permet de déterminer des sous-communautés de *reviewers* au sein des réseaux EDM et LAK, c'est-à-dire les ensembles de sommets fortement reliés entre eux. Selon la définition proposée par Newman [NEWMAN 2006], l'algorithme de modularité calcule le nombre de liens dans chaque groupe moins le nombre de liens dans les mêmes groupes, dans un graphe où les liens auraient été distribués de façon

aléatoire. Il en ressort une représentation similaire pour les deux communautés. Au cours des années de fondation (de 2008 à 2012 pour EDM, en 2011-2012 pour LAK), le vivier des *reviewers* est stable. Il connaît dans les deux communautés EDM et LAK un profond renouvellement en 2013-2014. Une troisième génération de *reviewers* apparaît en 2015-2016 tant chez EDM que LAK à cette différence près : le renouvellement est tel dans le *Program Committee* LAK que chacun de ces millésimes constitue une sous-communauté.

**Figure 1. Les reviewers des *Program Committees* des conférences EDM et LAK**



Ces graphes d'affiliation entre conférences et *reviewers* montrent-ils une hybridation avancée entre les deux communautés EDM et LAK ? Le nombre de *reviewers* présents à la fois aux *Program Committees* EDM et LAK reste modeste (Figure 1c). De 2011 à 2016, seuls 10% des *reviewers* ont été associés à ces deux conférences (33 sur 339). Au-delà des efforts menés par leurs états-majors en 2011 et 2013, ces deux communautés continuent donc d'entretenir de fortes spécificités. En témoignent la faible part des auteurs communément publiés dans les revues *JEDM* et *JLA* (9 auteurs sur 271, cf Tableau 13 en annexe).

### 1.3 Spécificités et objectifs des Learning Analytics

Il faut clairement distinguer les *Learning Analytics* des *Academic Analytics*. Proposer des méthodes, visualisations, algorithmes en vue d'améliorer les résultats des apprenants, renforcer leur engagement, optimiser leurs expériences d'apprentissage : tout ceci relève des *Learning Analytics*. Soutenir la représentation de minorités ethniques, augmenter la productivité de l'organisation, allouer des ressources aux établissements en déficit de résultats : ces objectifs relèvent en revanche de pilotages institutionnels et de stratégies politiques (vs SCLATER 2014, p. 4). Les *Learning Analytics* ont une visée cognitive. Appliquer les recettes de l'informatique décisionnelle (ou *Business intelligence*) à la sphère éducative en vue du pilotage institutionnel relève des *Academic Analytics* [CAMPBELL *et al.* 2007 ; LONG, SIEMENS 2011]. La confusion reste vivace mais abusive.

**Tableau 2. Niveaux, objets d'analyse et acteurs en Academic & en Learning Analytics d'après LONG, SIEMENS 2011, p. 4.**

Academic Analytics			Learning Analytics		
Niveaux	Analyses	Bénéficiaires	Niveaux	Analyses	Bénéficiaires
Local	profils d'apprenants, performances académiques, <i>knowledge flow</i> , concentration des moyens financiers	Administrateurs finances, marketing	individuel	compréhension de la performance personnelle en relation avec des objectifs d'apprentissage et habitude de travail des camarades.	Apprenants Personnels d'éducation et enseignants
Régional	comparaison entre systèmes éducatifs, qualités et standards, classements	Finances, gestionnaires, administrateurs	cohortes	réseaux sociaux, développement conceptuel, analyse de discours, progression intelligente	Enseignants
National & international		Gouvernements, organisations, UNESCO, OCDE	Équipe pédagogique	modélisation prédictive, modèles d'échec/succès	Personnels d'éducation et enseignant

Fondées sur la collecte de données pour chaque apprenant, les *Learning Analytics* permettent une communication différenciée aux acteurs de l'apprentissage en variant les niveaux d'analyse, le temps écoulé et la granularité des données [BROWN 2012]. Au fil des contributions, les *Learning Analytics* élargissent leurs méthodes à cinq champs d'action. Quels sont-ils ?

**Tableau 3. Des objectifs communs à EDM et aux Learning Analytics, d'après BAKER 2014, p. 257-262**

	OBJECTIFS DU TRAITEMENT	MÉTHODES	QUELQUES TRAVAUX DE RÉFÉRENCE
PREDICTION ENTRE VARIABLES	Développer un modèle permettant de prédire une variable (dépendante) à partir de la combinaison d'autres variables (indépendantes). A partir d'un jeu de données complet mais de taille réduite, on détermine la valeur de la variable à prédire. Le modèle est statistiquement validé de façon à être appliqué à plus grande échelle.	Classification (variable prédite binaire ou catégorielle) : arbre de décision, forêt d'arbres décisionnels, règles de décision, régression séquentielle et logistique. Régresseurs (variable continue). Estimation de connaissance latente (type spécifique de classifieurs)	Travaux de portée générale : DEKKER <i>et al.</i> 2009 ; FENG <i>et al.</i> 2009, MING, MING 2012.  Les algorithmes d'estimation de connaissance sont au fondement des tuteurs intelligents (KOEDINGER, CORBETT 2006)

DECOUVERTE DE STRUCTURES	A l'inverse d'une tentative de prédiction sur une variable dépendante, cette méthode vise à établir la structure des données sans idée préconçue sur l'objet recherché, sans intérêt a priori pour aucune variable.	Clustering ou partitionnement des données : trouver les données qui se regroupent naturellement entre elles, en séparant le jeu en plusieurs sous-ensembles	Grouper des étudiants (BEAL <i>et al.</i> 2006). Différencier les actions des apprenants (AMERSHI, CONATI 2009)
		Analyse factorielle : utilisée dans le même but de créer des sous-ensembles homogènes.	Quels choix de design sont généralement faits par les concepteurs des tuteurs intelligents ? (BAKER <i>et al.</i> 2009).
		Analyse de réseaux sociaux : des modèles sont développés à partir des relations et interactions entre individus.	Evaluer l'efficacité de différents groupes de projet (KAY <i>et al.</i> 2006). Positionner des apprenants dans le réseau et mesurer leur perception de la communauté éducative (DAWSON 2008) Evaluer l'engagement d'un apprenant dans sa scolarité (MACFADYEN, DAWSON 2010)
		Découverte du modèle du domaine	Tracer l'apprentissage durant l'utilisation d'un tuteur intelligent (CEN <i>et al.</i> 2006)
FOUILLE DE RELATIONS	Découvrir les relations entre variables au sein d'un jeu de données avec un grand nombre de variables, soit pour trouver quelles variables sont étroitement associées avec une tierce variable d'intérêt, soit quels couples de variables sont étroitement reliés.	Fouille de règles d'association	Trouver des patterns d'étudiants obtenant de bonnes performances pour pouvoir faire de meilleures suggestions aux étudiants qui éprouvent des difficultés : BEN-NAIM 2009.
		Fouille de corrélation : trouver des corrélations linéaires positives ou négatives entre variables.	Corrélations calculées parmi les variables sur la conception des leçons d'un système de tuteur intelligent et la prévalence des étudiants détournant le système (étudiants qui utilisent le logiciel pour avancer sans consulter la documentation du cours). De petits problèmes scénarisés mènent à une plus grande proportion d'étudiants détournant le système que des scénarisations riches ou pas de scénarisation du tout (BAKER <i>et al.</i> 2009)
TRAITEMENT DE DONNEES POUR EVALUATION HUMAINE	Représenter les données d'apprentissage de façon efficace permet d'agir sur la pédagogie.	Méthodes de visualisations	Aide à la prise de décision : visualisation de la trajectoire des étudiants durant leur scolarité : identification de schémas parmi les étudiants ayant réussi ou non ; inférence d'étudiants à risques suffisamment tôt pour guider des interventions (BOWERS 2010).
		Carte de chaleur ou nuages de points, courbes d'apprentissage, diagrammes	Evolution des performances KOEDINGER <i>et al.</i> 2010, mesure de la motivation HERSHKOVITZ, NACHMIAS 2008
DECOUVERTE AVEC LES MODELES	Les résultats d'une analyse par fouille de données sont utilisés dans une autre analyse par fouille de données.  Les modèles utilisés ne sont pas nécessairement obtenus par des méthodes de prédiction mais ils peuvent être obtenus à travers d'autres approches comme le <i>clustering</i> , ou l'ingénierie des connaissances (STUDER 1998).	Modèles de prédiction utilisés au sein d'un autre modèle de prédiction	Les modèles de prédiction d'apprentissage robuste des étudiants (BAKER <i>et al.</i> 2011) ont généralement dépendu de modèles sur les comportements métacognitifs des étudiants (ALEVEN <i>et al.</i> 2006), qui à leur tour dépendent d'évaluations de connaissance latente d'étudiants (CORBETT, ANDERSON 1995) qui ont dépendu de modèles sur la structure du domaine (cf. KOEDINGER <i>et al.</i> 2012).
		Un modèle de prédiction est utilisé dans une analyse de fouille de relations : à l'étude, la relation entre les prédictions du modèle initial et des variables additionnelles	BEAL <i>et al.</i> 2008 : développe un modèle de prédiction d'étudiants détournant un système et le corréle aux différences individuelles entre étudiants pour comprendre quels étudiants sont plus vraisemblablement engagés dans ce type de comportement.

## 2 Questions de recherche en science des données de l'apprentissage

Les tenants des *Learning Analytics* partent du principe que la numérisation des activités d'apprentissage va offrir aux acteurs de la chaîne éducative des données robustes, issues du terrain. Qu'ils mobilisent ces indicateurs en présentiel comme à distance, en temps réel comme en différé, à un niveau personnel ou agrégé, ces acteurs seraient constamment en position de faire des choix éclairés pour améliorer leur expérience.

C'est à présent à un panorama de ces recherches expérimentales que nous convions le lecteur. Cet état de l'art ne peut rendre compte de l'ensemble des travaux entrepris dans le domaine des sciences des données de l'apprentissage mais il prétend éclairer les questionnements les plus actifs et les recherches les plus abouties. A l'hétérogénéité du millier d'articles imprimés dans les *Proceedings* des Conférences, nous avons privilégié une revue exhaustive des articles passés par le second tamis du *Journal of Educational Data Mining* (JEDM) et du *Journal of Learning Analytics* (JLA), depuis leur création. Nous présentons ainsi 17 titres récemment parus, sous forme de tableaux, sélectionnant les travaux dont l'objectif, les techniques et les résultats nous semblaient les plus significatifs.

Pour chacun des thèmes retenus, ces recherches peuvent être motivées par des logiques diverses : d'une part des protocoles de recherche fondamentale ambitionnent de se rapprocher au plus près de la boîte noire en captant la moindre interaction de l'apprenant pour agir sur son apprentissage et l'améliorer. D'autre part, une recherche davantage appliquée se soucie de donner aux acteurs institutionnels les moyens de mesurer les flux d'apprenants et d'agir sur les performances du système, sur les succès et les échecs observés de l'apprentissage, en particulier l'abandon (*drop out*).

## 2.1 Prédire la progression de l'apprenant

La prédiction du parcours (comportemental ou cognitif) d'un apprenant est l'un des plus anciens problèmes des Environnements Informatiques pour l'Apprentissage Humain (EIAH), des systèmes experts aux MOOCs. Avec l'acquisition en temps réel de données d'apprentissage toujours plus fines et massives, les chercheurs ont développé des techniques d'analyse issues du *data mining* pour classer des profils, les orienter en fonction des capacités estimées, adapter les contenus, déployer des stratégies d'engagement et lutter contre le décrochage. Ces analyses sont menées à différents niveaux de granularité, de la simple interaction en temps réel d'un individu (microgenèse) aux apprentissages d'une cohorte sur une période donnée. Dans les deux communautés EDM et SoLAR, des modèles prédictifs et des méthodes variées sont mises en œuvre comme l'atteste le tableau suivant (Tableau 4).

**Tableau 4. Objectifs, techniques et apports des modèles de prédiction**

Référence	Objectif	Données	Techniques	Résultats
ZIMMERMANN <i>et al.</i> 2015	Évaluer la puissance de prédiction des résultats scolaires et leur agrégation, comme indicateurs de performance	81 variables pour une population de 171 étudiants	Post hoc. Modèle de régression	Les résultats de licence ( <i>undergraduate</i> ) peuvent expliquer 54% de la variance dans les résultats de cycles supérieurs. La moyenne de notation globale de la 3 <sup>e</sup> année est la variable la plus significative. Les résultats fournissent une base méthodologique pour dresser des lignes générales pour les comités d'admissions.
KNOWLES 2015	Dans 1000 écoles du Wisconsin, un système évalue la probabilité de passage pour chacun des 225 000 collégiens ( <i>grade 6 to 9</i> )	Cohorte des 12-13 ans en 2005. Les variables portent sur les résultats, l'assiduité, le comportement, la mobilité entre écoles, etc.	Deux fois par an. Régression logistique	Le système fournit une probabilité de passage pour chaque apprenant + un classement (bas, modéré, haut). Les étudiants reçoivent une catégorie de risque pour 4 sous domaines : scolarité, présence, comportement, mobilité. Le système identifie 65% des échecs et des retards dans l'avancement avant l'entrée au lycée ( <i>high school</i> ) avec de faibles taux de fausses alarmes.
FERGUSON, CLOW 2015	Évaluent la méthode de <i>clustering</i> mise au point par Kizilcek pour déterminer les modèles d'engagement sur un MOOC.	24 000 participants à 4 MOOCs délivrés par <i>The Open University</i> , en Sciences Physiques, Sciences de la Vie, en Arts et en Economie	<i>Clustering</i> (partitionnement de données)	KIZILCEK <i>et al.</i> 2013 établissent une méthode de <i>clustering</i> permettant d'identifier 4 modèles d'engagement : les apprenants qui rendent tous les devoirs, ceux qui voient les vidéos mais ne rendent pas tous les devoirs, ceux qui abandonnent en cours de route et ceux qui explorent quelques vidéos. Appliquée au nouveau jeu de données, ni la méthode ni les modèles ne semblent adaptés à une ingénierie pédagogique sensiblement différente de celle du MOOC étudié par Kizilcek. L'enjeu est alors de proposer une nouvelle méthode d'analyse temporelle.
AGUIAR <i>et al.</i> 2014	Prédire l'attrition grâce à la mesure de l'engagement, fondée sur l'usage du portfolio numérique.	Portfolios électroniques de 429 étudiants en 1 <sup>ère</sup> année d'Ingénierie à Notre Dame [IN-US]	Algorithmes de classification (Naïve Bayes, arbre de décision, régression logistique...)	Les auteurs proposent une méthode pour mesurer l'engagement des étudiants à partir de leurs portfolios numériques et montrent en quoi ces nouveaux indicateurs peuvent améliorer la qualité de prédiction.
MARTIN <i>et al.</i> 2013	Établir les différents chemins dans l'apprentissage des fractions avec le jeu en ligne : <i>Refraction</i>	Codage <i>a posteriori</i> des transitions entre états du jeu.	Fouille de données et visualisations de graphes	Un algorithme de classification permet de regrouper les étudiants en fonction des transitions entre états du jeu. Ils identifient des profils d'apprenants : ceux allant droit à la solution, ceux qui expérimentent.

La problématique de la prédiction sert deux logiques différentes. Au niveau macro, dans une logique de recherche appliquée, les institutions éducatives sont en quête de modèles pour prévenir l'échec et améliorer leurs résultats. Au niveau micro, dans une perspective plus expérimentale, les chercheurs tentent d'inventer de nouveaux

modèles pour se rapprocher au plus près de la boîte noire de l'apprenant en captant la moindre interaction dans l'espace d'une situation d'apprentissage : une frappe, un clic, un coup d'oeil sont autant de traces pour informer les processus de l'apprentissage. Dans un cas comme dans l'autre, des modèles préexistants à la naissance des communautés EDM et LAK sont repris, itérativement affinés et adaptés aux nouveaux environnements d'apprentissage et à leurs enjeux.

Ainsi, de nombreuses solutions commerciales fondées sur différents modèles de prédiction du décrochage s'épanouissent sur les marchés éducatifs les plus concurrentiels. John P. Campbell implémente en 2007 sur le LMS *Course Signals* un calcul de régression logistique fondé sur les résultats, l'engagement, les indicateurs sociologiques, le nombre de messages postés sur le forum, de messages envoyés et de devoirs complétés [ARNOLD, PISTILLI 2012]. A. Essa et H. Ayad développent en retour, pour le LMS *Desire2Learn*, le *Student Success System*, un programme alliant une plus grande diversité de modèles à un outil de diagnostic.

Les données relatives aux activités d'apprentissages, aux pré-requis, à l'ingénierie pédagogique et à la dynamique du cours, aux modalités d'apprentissage (en ligne, en présentiel ou hybride) sont mobilisées selon une approche comportementale. De la même façon que la présence en ligne d'un apprenant est évaluée selon ce même calcul de régression logistique, sa capacité à achever un parcours, sa participation et sa sociabilité sont évaluées. Le gestionnaire du cours choisit un ensemble de règles pertinentes, et leur poids relatif, pour obtenir une probabilité selon trois niveaux : "à risque", "risque potentiel" et "succès" [ESSA, AYAD 2012]. Ces recherches se poursuivent avec la massification de la formation en ligne et le succès des MOOCs à partir de 2012. Doug Clow quantifie l'usure de la motivation d'un apprenant *via* ses activités sur 3 sites en ligne [CLOW 2013]. Kizilcec présente une méthode de classification qui identifie différentes trajectoires d'engagement basées sur les interactions de l'apprenant avec les lectures de vidéos et les exercices [KIZILCEC *et al.* 2013]. Première tentative de déconstruction du désengagement, ce papier est destiné à une constante réévaluation [FERGUSON, CLOW 2015].

Dans la microgenèse de l'apprentissage, les trois modèles diagnostiques au fondement de la recherche expérimentale, ont été posés dans la décennie 90, se sont considérablement développés au bénéfice de la multiplication des EIAH tels les tuteurs intelligents, mais peinent encore à être adaptés au format des LMS et MOOCs. Le modèle de *Knowledge tracing* [CORBETT, ANDERSON 1995] dérive de la théorie ACT-R, qui vise à modéliser l'ensemble des processus cognitifs [ANDERSON 1983]. Postulant que l'apprentissage de nouvelles connaissances est procédural, ce diagnostic vérifie le transfert de la connaissance de la mémoire déclarative à la mémoire procédurale. A ce modèle s'ajoute un second : le diagnostic *Constraint-based modeling* [MITROVIC *et al.* 2007, OHLSSON 1994] qui stipule que l'apprentissage se fait par la confrontation de l'apprenant avec ses erreurs et se base sur la théorie *performance errors* [OHLSSON 1996]. Seuls les états du problème ayant un intérêt pédagogique sont modélisés, permettant de relever une erreur typique du domaine. Enfin le diagnostic *Control-based* [MINH CHIEU *et al.* 2010] est basé sur le modèle didactique cKc [BALACHEFF, GAUDIN 2002; BALACHEFF 1995] implémenté comme un réseau bayésien dynamique.

Des évolutions de ces modèles sont ainsi proposées pour passer d'une démarche centrée expert vers une démarche centrée données comme avec le modèle de régression linéaire *Additive Factor Model* ou AFM [CEN *et al.* 2008] et les évolutions *Performance Factor Model* ou PFM [PAVLIC *et al.* 2009] ou *Performance Factor Analysis* (PFA). Enfin, ces approches centrées données permettent également la comparaison entre les modèles d'apprenant de façon plus systématique tout en maintenant la possibilité de la répliquabilité des résultats [LALLE *et al.* 2013].



## 2.2 Mesurer les interactions sociales

En s'appuyant sur les apports fondamentaux de Vitgovsky et du socio-constructivisme, une communauté scientifique s'attache à évaluer le rôle des interactions sociales au sein des EIAH : la communauté *Computer-supported collaborative learning* (CSCL), née en 1995, porte cette conviction que la cognition et l'action humaine sont socialement et culturellement médiées. L'objectif est donc de comprendre et valoriser ces processus d'apprentissage collectif à l'encontre d'un enseignement transmissif, dont le Graal demeure la performance individuelle dans le cadre d'évaluations sommatives. L'apprentissage social permet aux individus de clarifier leurs intentions, étayer leur apprentissage, approfondir par l'échange. Son modèle émane des interactions de trois dimensions : cognitive (la situation problème), affective (la motivation) et sociale (les échanges).

L'analyse de l'apprentissage social (*Social Learning Analytics*) s'appuie notamment sur l'analyse des réseaux sociaux, permettant de détecter des communautés [CLAUSET *et al.* 2004, FORTUNATO 2010] ; d'identifier des sous-ensembles cohésifs dans un réseau (mesurable en terme de proximité, fréquence, affinités) [REFFAY, CHANIER 2003] ; d'investiguer sur la densité de ces réseaux [BORGATTI *et al.* 2009] ; et dans le cas d'un réseau égocentrique, d'identifier les personnes aidantes ou avec lesquelles surgissent des conflits liés à une incompréhension mutuelle [HAYTHORNWAITE, DE LAAT 2010]. Caractériser les liens entre acteurs ajoute une dimension à l'analyse : les individus se fient généralement à tous types d'interlocuteurs quand il s'agit d'accéder à de nouvelles connaissances ou d'apprendre de façon informelle ; mais ils se tournent vers des personnes de confiance pour approfondir des connaissances [LEVIN, CROSS 2004]. Le concept de *Social Learning Analytics* émerge autour, notamment, des travaux de Ferguson [FERGUSON 2009]. L'analyse de l'apprentissage social tend ainsi à révéler les régimes d'engagement des apprenants, soit dans une activité sociale (envoyer un message, sympathiser, suivre un tiers), soit en créant des traces réutilisables (publier, rechercher, taguer, évaluer).

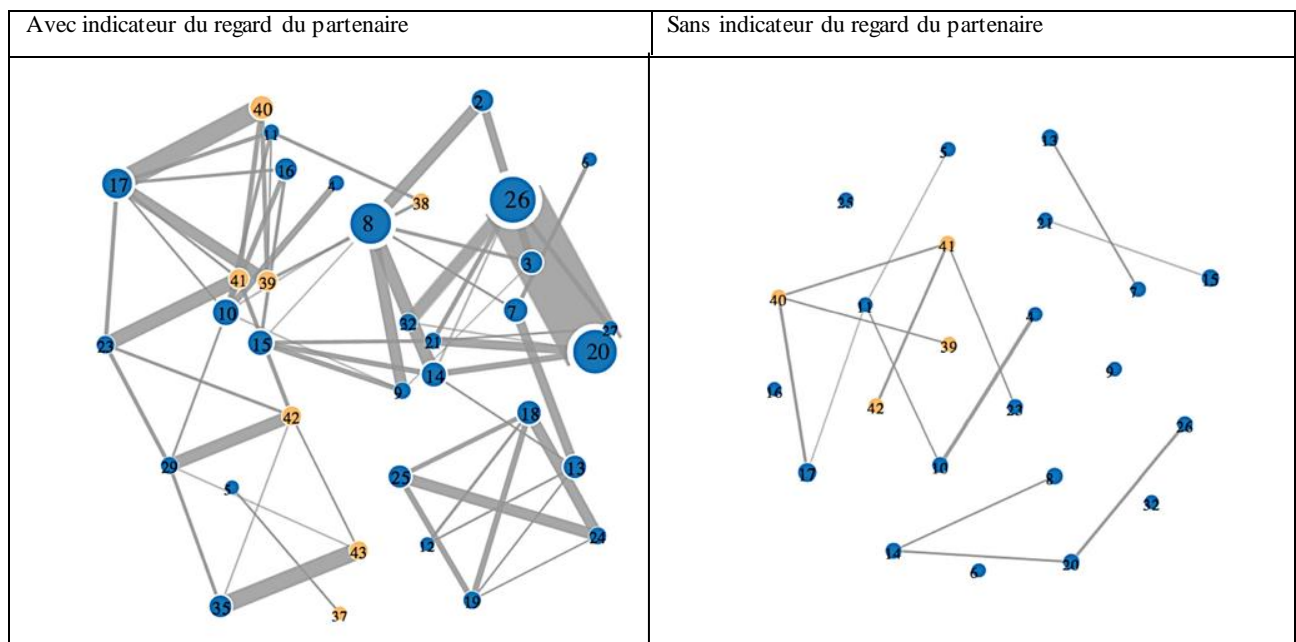
**Tableau 5. L'actualité de la recherche en *Social Learning Analytics***

Référence	Objectif	Données	Techniques	Résultats
JOKSIMOVIC <i>et al.</i> 2015	Mesurer, dans un cMooC, l'influence des modèles de discours sur le capital social des apprenants	17324 messages sur deux sessions du MooC <i>Connectivism and Connective Knowledge</i>	Graphes non dirigés hebdomadaires pour chaque média (FB, Twitter, blogs). Analyse linguistique selon la métrique COH	La qualité des écrits est déterminante pour l'apprentissage. Les apprenants d'un cMooC se joignent plus volontiers à des individus qui utilisent un style informel, narratif et cependant cohérent. Le langage définit donc la position structurelle dans un réseau social d'apprentissage.
WATERS <i>et al.</i> 2014	Identifier, classifier les formes de collaborations (fraude, compétences individuelles)	Données fictives et réelles : examen de 203 vrai/faux passé par 97 étudiants d'informatique	Approche bayésienne, Modèle de Rash, méthode <i>Sparse Factor analysis</i>	A partir de deux nouveaux algorithmes, les auteurs ont su, tant sur des données fictives que réelles, détecter et identifier des formes de collaboration. Cette méthode permet à l'instructeur de se focaliser sur un petit nombre d'apprenants.
SKRYPNYK 2014	Analyser le réseau social d'un MooC connectiviste à travers les interactions via Twitter	2483 tweets de 800 participants. Données démographiques publiques sur les réseaux et le web	Mesures au niveau des nœuds (centralité de degré, de proximité, d'intermédiarité, de vecteur propre, de clique) et du réseau (modularité)	Au cours de l'avancement du MooC un groupe de participants a joué un rôle similaire à celui des animateurs. Les nœuds avec les plus hauts degrés de centralité concernent les <i>hashtags</i> , confirmant ainsi l'intérêt des apprenants pour une des thématiques changeantes. Des sous-ensembles ont émergé avec, pour chacun, des contributions spécifiques.
SCHNEIDER <i>et al.</i> 2013	Tester une analyse séquentielle des regards et évaluer la robustesse de cette métrique pour	22 dyades isolées communiquent via un canal audio seul ou un canal audio + un	La position du regard (nœuds) et son mouvement (transition) sur un support pédagogique est	Les auteurs démontrent l'importance de considérer le regard comme indice d'attention soutenue mutuelle. En adaptant la théorie des graphes à leur problématique, ils identifient de nouveaux indicateurs (nombre de liens, taille des

	prédire l'apprentissage	indicateur de regard du collaborateur	transformé en graphes non dirigés	nœuds, intermédiarité) permettant d'évaluer l'intensité d'une collaboration et la compréhension réciproque des apprenants.
AHN 2013	Identifier des types de comportement, mesurer la corrélation entre ces comportements et les compétences du 21 <sup>e</sup> siècle	Enquête sur 189 participants. Collecte de données agrégées via l'API Facebook de 99 apprenants	Analyse factorielle, modèles de régression	Le modèle de régression prend les activités pour variables indépendantes (envoyer un message, partager une information, devenir ami, rejoindre un groupe) et les compétences pour variables dépendantes (négociation, réseautage, esprit critique, jeu, multitâche, appropriation, navigation transmedia). Il illustre quel type d'activités renforce quelles compétences.

Ce domaine de recherche prend de l'ampleur avec la diffusion massive d'un nouveau type de formation en ligne à compter de 2011 : les MOOC (*Massive Open Online Courses*). Des cohortes de dizaines de milliers d'apprenants se retrouvent sur des plateformes d'apprentissage en ligne pour des formations relativement courtes (de 6 à 12 semaines) nécessitant dans le cas des Mooc connectivistes (les cMooc) de travailler de façon collaborative. C'est l'échelle même des expérimentations qui s'en trouve bousculée : on passe dès lors d'échantillons relativement restreints à la mobilisation de gisements de données considérables. Si les bases théoriques de cette analyse des interactions sociales sont largement fondées par les recherches des années 1990-2000, les emprunts plus récents à l'analyse des réseaux sociaux permettent d'élargir les techniques de mesures et de représentations de nouvelles biométriques, dont tout porte à croire qu'elles vont être toujours plus précises : l'usage d'une caméra *kinect* permet déjà d'étudier la synchronisation des postures physiques, et pourra demain être corrélé à de nouveaux indicateurs sur la voix ou le rythme cardiaque, pour *in fine* révéler l'intensité des interactions sociales.

**Figure 2. Graphes d'analyse de l'activité collaborative**



Le réseau est constitué de 44 nœuds : ces nœuds signalent la topologie d'un support pédagogique remis aux étudiants pour apprendre le traitement de l'information visuelle par le cerveau humain. De taille variable, nœuds et liens indiquent les regards partagés au sein d'une dyade d'apprenants. Les graphes avec les plus denses lieux/position et transitions de regards partagés indiquent une mutuelle orientation pour accomplir la tâche.

Le graphe à droite montre un groupe qui travaille sans indicateur du regard du partenaire. La collaboration est pauvre : les étudiants sont rarement concentrés sur la même partie du document, les nœuds sont petits et très peu reliés. Le graphe à gauche montre un groupe bénéficiant de cet indicateur du regard d'autrui : une attention soutenue et simultanée contribue à l'approfondissement de l'apprentissage par un échange de plus grande qualité. Les dimensions du graphe informent donc la qualité de l'apprentissage collaboratif : la taille d'ensemble permet de mesurer la réciprocity d'un apprentissage ; la taille des nœuds permet d'évaluer à quel point une dyade a travaillé à dégager un consensus ; enfin la centralité intermédiaire (mesure de centralité d'un sommet au sein d'un graphe) permettrait de comprendre à quel point des étudiants se comprennent [SCHNEIDER *et al.* 2013].

## 2.3 Analyser le discours

L'apprentissage ne résulte pas des seules capacités cognitives ou du comportement d'un apprenant isolé [MERCER 2004]. Le langage est l'un des premiers vecteurs par lequel les apprenants construisent du sens : son usage est influencé par les buts, les sentiments et les relations, très variables selon les contextes [WELLS, CLAXTON 2002]. Le langage fait partie de la situation d'apprentissage [GEE, GREEN 1998, WERTSCH 1991]. Il supporte et conditionne le succès d'activités conjointes tant pour la combinaison des connaissances, des compétences, l'utilisation d'outils et la capacité à travailler ensemble. Comprendre l'apprentissage de cette façon requiert de faire attention aux processus de construction du savoir dans le groupe. L'analyse du discours tend à rendre ces processus d'apprentissage visibles et ouvre la possibilité de les améliorer à différents niveaux, de l'individu au groupe restreint, puis à la cohorte.

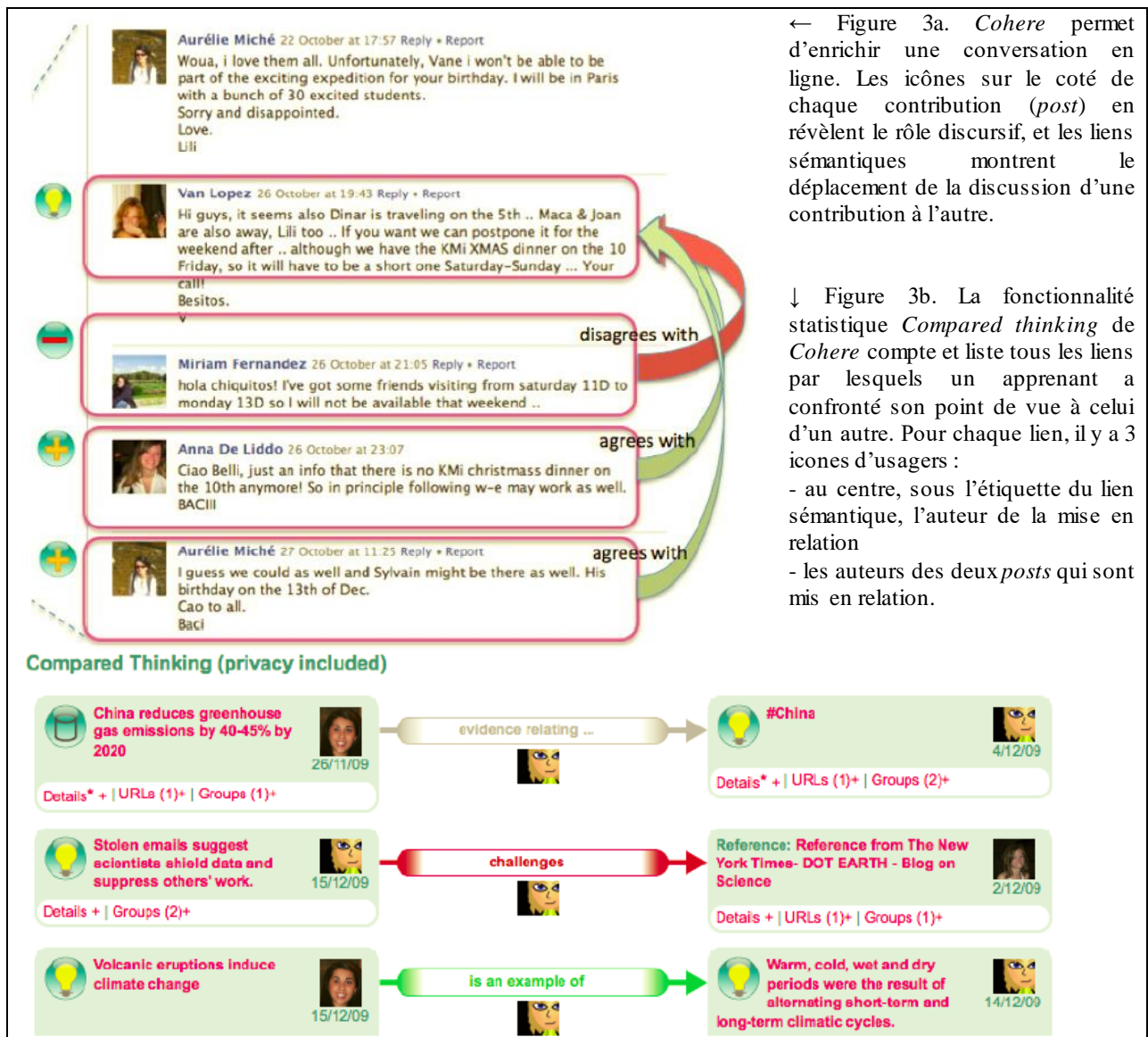
Avec le déploiement de l'apprentissage à distance, le forum, lieu d'une discussion asynchrone, devient un objet de recherche. Les forums aident à trouver et partager l'information et promeuvent la réflexion critique de communautés d'apprenants [ANDERSON 1996]. Plusieurs modèles d'apprentissage en ligne ont été développés dont celui de *Communauté d'interrogation (Community of Inquiry, CoI)* [GARRISON *et al.* 1999]. Ce modèle part du principe que dans toute collectivité, l'apprentissage se fonde sur l'interaction de trois éléments principaux : la présence de la société signifie la capacité des apprenants à participer au plan social et affectif à la vie de la collectivité ; la présence de la cognition indique l'aptitude à fabriquer du sens grâce à une communication soutenue des milieux en ligne ; la présence de l'enseignement désigne des activités visant à faire activement participer les apprenants et à maintenir la communication. Analyser une conversation entre apprenants permet d'appréhender leur degré d'ouverture, leur capacité de décentrement et leur propension à reformuler leur pensée. Trois modalités distinguées par Mercer en : discours dialectique, cumulatif et exploratoire [MERCER 2004].

**Tableau 6. L'actualité de la recherche en Analyse de discours**

Référence	Objectif	Données	Technique	Résultats
SNOW <i>et al.</i> 2015	Proposer une méthodologie pour valuer la souplesse stylistique des rédacteurs confirmés	45 lycéens complètent 2 tests (capacités avant/après) et 16 essais en 8 sessions	Traitement automatique du langage naturel (TALN) (narrativité, cohésion), entropie	Les auteurs proposent une méthode d'évaluation discrète et rapide permettant de détecter les relations entre souplesse stylistique, lecture, vocabulaire, connaissances et écriture. Seule la combinaison entre TALN et entropie permet de saisir la progression des apprenants vers plus de souplesse stylistique.
REBOLL EDO-MENDEZ <i>et al.</i> 2014	Dispositif pour rechercher des modèles d'interaction en lien avec l'étayage motivationnel d'un tuteur intelligent	70 fichiers de logs capturant le comportement de 35 apprenants avec un tuteur intelligent	Post hoc. Corrélation entre types collaboratifs. Mesure d'association entre les actions de l'apprenant. Cf D'Mellow 2010	Un graphe dirigé permet de visualiser et d'identifier des transitions significatives dérivées de dyades d'actions. Première approche à compléter par des entretiens avec les apprenants.
MING MING 2014	Identifier le lien entre cognition informelle (opinion, anecdotes, description) et formelle (explication théorique)	1330 messages asynchrones écrits par 17 étudiants, codés par eux-mêmes, sur un cours en ligne	Méthode de Monte Carlo par chaînes de Markov (MCMC) itérative	Evalue comment 3 types cognitifs (opinion informelle, élaboration, preuve) et 3 types de métacognition sociale (demander une explication, demander comment utiliser, opinions différentes) renforcent la vraisemblance de nouvelles informations ou explications théoriques dans les messages postérieurs.
D'MELLOW <i>et al.</i> 2010	Méthode pour la détection automatique de modèles collaboratifs avec un tuteur intelligent	50 heures de tutorat filmées entre apprenants et experts et 47 296 interactions codées	Construction d'un graphe orienté des transitions entre interactions de l'apprenant et test d'hypothèses	A la différence des modèles de Markov cachés, des règles d'association séquentielle et des modèles log-linéaires, cette étude détecte les transitions fréquentes et les représente sous forme de graphe. L'étude des stratégies, actions et dialogues d'experts humains permet aux auteurs de concevoir la modélisation informatique d'un tuteur intelligent.

Dans la lignée de ces apports théoriques, ce sont les communautés investies dans la conception de modèles de remédiation au sein d'un tuteur intelligent qui se sont employées à dégager des indicateurs permettant d'identifier automatiquement les modalités d'interventions pour étayer un apprentissage. Avec l'émergence des réseaux sociaux, les chercheurs en *Learning analytics* s'efforcent d'identifier les types d'échanges : challenges, explorations, évaluations, raisonnement. Ainsi, la plateforme de délibération *Cohere* a été étendue par De Liddo et ses collègues pour fournir des *Learning Analytics* qui identifient : les sujets de conversation et le point de vue des apprenants ; les types de contributions des apprenants et leurs accords/désaccords (Figure 3a) ; l'organisation du réseau discursif et le rôle des apprenants dans ce réseau (propositions/discussions, Figure 3b) ; les liens sémantiques entre apprenants [DE LIDDO *et al.* 2011].

**Figure 3. Du marquage à l'analyse statistique de l'activité collaborative sur *Cohere***



## 2.4 Donner à voir l'apprentissage

A l'origine des *Learning Analytics*, les *Visual Data Analytics* reposent sur deux principes essentiels : donner aux acteurs de la communauté éducative un pouvoir de décision par l'exploration du modèle qui sous-tend les données d'apprentissage elles-mêmes [DUVAL 2011]. Les tableaux de bord (*Dashboards*) tiennent des systèmes d'aide à la décision (*Decision Support System*, DSS) apparus dans les années 70, dans la foulée de la numérisation des *process* métiers. Dans le domaine éducatif, les tableaux de bords numériques se sont imposés avec les premiers LMS (*Learning Management Systems*). Tous les LMS ne proposent pas des interfaces de *Visual Data Analytics* aussi poussée que *Blackboard Analytics*, *Brightspace*, *Signals* (Purdue University) et *ALAS-KA* (Khan Academy). Ces logiciels prétendent au-delà d'une consultation des données d'apprentissage, permettre de dépister le décrochage, de personnaliser les contenus. En permettant aux apprenants d'accéder à leurs propres données, ces technologies coïncident avec l'émergence du *Quantified Self* ou *Self Tracking* (mesure de soi) : des techniques d'évaluation quantitative systématique.

Les tableaux de bord numériques (*Dashboards*) offrent une interprétation visuelle de larges ensembles de données pour, en quatre étapes, découvrir, interroger, comprendre les modèles portés par ces données et, *in fine*, modifier ses représentations [VERBERT *et al.* 2013]. Ces tableaux visent à seconder l'enseignant dans le développement de l'attention, la compréhension et la métacognition des apprenants. Ils permettent d'un simple clic de passer d'une visualisation de haut-niveau à l'interrogation des données de bas-niveau. Pour répondre aux aspirations d'une économie de l'attention [GOLDHABER 1997], la conception de l'interface (*User Interface Design*) devrait être claire, interactive et modulable de façon à pouvoir répondre aux besoins spécifiques de chaque type d'utilisateurs. Les tableaux de bord sur le marché sont relativement bien documentés sur le plan scientifique [VERBERT *et al.* 2013]. Santos et al. présentent les résultats du tableau de bord *StepUp!* et discutent son apport aux problèmes et besoins des apprenants [SANTOS *et al.* 2013]. Les 1500 étudiants de Purdue ayant expérimenté le tableau de bord *Course Signal* ont obtenu des résultats significatifs d'assiduité, en comparaison avec des cohortes similaires [ARNOLD, PISTILLI 2012].

Au département de Sciences de l'Éducation à Utrecht (Pays-Bas), van Leeuwen et ses collègues ont évalué la pertinence du tableau de bord dans la gestion de classe. Un enseignant doit surveiller les activités cognitives de ses apprenants répartis en îlots [VAN LEEUWEN 2015]. Dans une première expérimentation (*The Concept Trail*), les apprenants doivent mener une tâche collaborative par messagerie synchrone. Au-fur-et-à mesure de la discussion, des concepts essentiels pour résoudre la tâche affluent. L'enseignant suit, en temps réel, la progression des groupes par le biais de lignes de temps où s'affichent ces concepts, échangés via la messagerie. Une deuxième expérimentation (*Progress Statistics*) permet à l'enseignant de suivre le nombre de mots écrits par les groupes d'apprenants dans l'éditeur collaboratif et dans le chat associé. Ces outils confèrent à l'enseignant une vue d'ensemble et une capacité d'intervention à distance qu'il ne peut avoir pendant un cours en regardant par dessus l'épaule de l'un ou de l'autre. Dans ces deux expériences, la visualisation n'améliore ni ne baisse la détection des problèmes cognitifs. En revanche elle augmente la fréquence des interventions de l'enseignant et renforcerait sa capacité à poser un diagnostic.

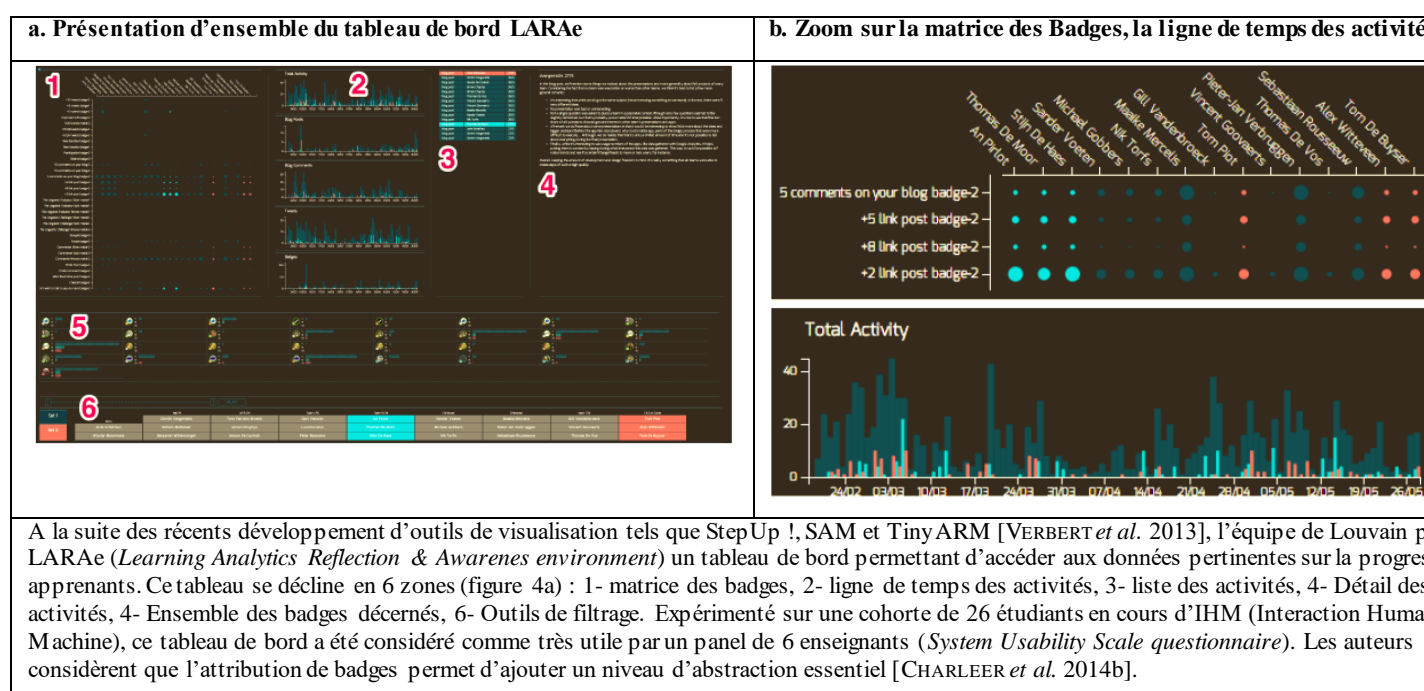
**Tableau 7. Evaluer l'utilité des Tableaux de bord pour l'apprentissage**

Référence	Objectif	Données	Dispositif	Apports et limites
VAN LEEUWEN	Assister en temps réel les enseignants dans l'évaluation	Évaluation de 5 groupes par 14 enseignants. Fichiers	Représentation de matrices avec en colonne les groupes	Cette étude permet de comprendre comment les enseignants interagissent avec des

2015	d'apprenants engagés sur une tâche collaborative	de log des actions de ces enseignants en temps réel	et en ligne la <i>timeline</i> des actions de l'enseignant	visualisations de données d'apprentissage.
MARTINEZ-MALDONADO <i>et al.</i> 2015	Elaborer un processus itératif de conception, validation et déploiement d'outils de visualisations pour favoriser l'attention de l'enseignant sur la collaboration et la progression des groupes	Manipulations physiques d'objets virtuels sur des tables interactives, temps de parole et progression des tâches de 500 étudiants	Entretiens et scénarii d'utilisation à partir de prototypes papier ; études contrôlées sur un simulateur ; analyse des interventions en situation réelle	Cette étude identifie 5 étapes dans le processus de conception d'outils de visualisation pour renforcer l'attention dans une activité de groupes sur tables interactives : identification du problème, validation des prototypes, simulation avec de vrais enseignants, études pilotes, expérimentation en classe. Des travaux complémentaires sont nécessaires pour adapter ce processus à un environnement d'apprentissage hybride ou à distance.
CHARLEER <i>et al.</i> 2014	Évaluer LARAE, un tableau de bord pour enseignants : visualisation des traces, badges, contenus des étudiants.	26 étudiants en groupes de 3. Matériaux et échanges en ligne. Évaluation l'utilisation de LARAE par 6 enseignants sur écran 27 pouces	Twitter API, fils RSS, Badge API. Visualisations : matrice des badges, ligne de temps, liste d'activités, détail des activités, options de filtrage	LARAE donne les moyens de visualiser l'abondance des traces laissées par les étudiants et les enseignants durant un cours. L'attribution de badges met l'accent sur les activités les plus importantes et les objectifs de formation. LARAE aide les enseignants et les étudiants à tirer parti des données qu'ils génèrent.

Sur un marché de plus de 2 milliards de dollars [ROBERGE 2013], les éditeurs de tableaux de bord s'engagent dans une levée de fonds et une bataille juridique intense. En réaction à l'accumulation de brevets, George Siemens porte, dès l'origine de SoLAR, un projet de conception d'une plateforme ouverte qui intégrerait une gamme variée de modules d'analyse dédiés à l'apprentissage [SIEMENS *et al.* 2011]. L'enjeu du libre répond alors à trois principaux défis : l'innovation et la généralisation des tableaux de bords exige que les traitements, les algorithmes et les technologies employés soient ouverts ; la plateforme doit être modulaire et permettre d'intégrer tous types d'outils liés à l'analyse et l'adaptation de l'apprentissage ; enfin, cette plateforme doit offrir des fonctionnalités appropriées à chacun, chercheurs et producteurs de contenus, assistants d'éducation et apprenants, enseignants et personnels d'encadrement. Si en 2016 ce projet ne semble pas avoir abouti, une autre initiative, APEREO *OpenDashboard*, est désormais accessible en Open Source.

**Figure 4. LARAE, un tableau de bord pour maîtriser les données d'apprentissage**



### 3 Les enjeux éthiques et déontologiques des *Learning Analytics*

Les quatre questions de recherche précédemment exposées ont solidement établi la plus-value du champ scientifique des *Learning Analytics* et fixé leur visée ultime : émanciper chaque acteur en décideur, adaptant ainsi la notion d'*empowerment*<sup>1</sup> au monde de l'éducation. Les dispositifs expérimentaux se diversifient qui soutiennent la mutation de l'apprentissage : l'instrumentation de la relation entre enseignant et enseignés permet une proximité croissante, un suivi plus personnalisé et l'émergence d'une communauté d'apprentissage. Pourtant un défi majeur reste de passer de protocoles expérimentaux à la diffusion de solutions innovantes. La montée en charge du numérique à l'Ecole constitue un enjeu financier, idéologique et éthique. Il s'agit donc dans cette dernière partie d'évoquer ce qui, dans la nature des changements en train de s'opérer, fait débat dans les communautés scientifique et éducative.

Les *Learning Analytics* consistent à fouiller des données d'apprentissage pour améliorer l'expérience éducative. Or, le volume de ces données est prêt à renseigner sur les moindres faits et gestes des individus, grâce à des capteurs et objets connectés, bien au-delà des simples données éducatives (lire une vidéo, répondre à un quiz). Jusqu'à quel point et comment mobiliser ces données, parfois si pertinentes, dans la perspective d'un projet éducatif ? Cette première question presse d'établir des contextes d'observation, des indicateurs, des processus d'analyse ouverts et discutables : c'est le projet d'HUBBLE en France.

Appliquées à l'Ecole, du primaire à la fin du secondaire, les *Learning Analytics* ont affaire à un public fragile et des données cognitives sensibles. Alors qu'un intense lobbying s'organise contre les législations européennes pour l'aliénation des données personnelles, comment le G29 (*Article 29 Working Party*, ensemble des CNIL européennes) envisage-t-il de protéger les usagers mineurs ?

De quelle Ecole numérique rêvons-nous à l'horizon 2025 ? Entre désirs et réalisme, huit *scenarii* sont soumis par le réseau LACE à la communauté des *Learning Analytics*.

---

<sup>1</sup> Le terme *empowerment* signifie littéralement 'renforcer ou acquérir du pouvoir'. Cette méthode d'éducation active vise à aider les acteurs de la communauté éducative à prendre conscience de leur rôle et d'acquérir les instruments qui leur permettront de faire des choix éclairés. Il est conçu comme une stratégie opposée au modèle dominant de transmission *top down*.

### 3.1 Sur le marché éducatif, le mouvement open-source

Jusqu'à présent les liens entre acteurs du marché éducatif restent relativement lâches. C'est l'édition des logiciels qui œuvre depuis une vingtaine d'années à mettre au point des plateformes de production et de collecte des données, des entrepôts (les *Learning Record Systems*), des tutoriels intelligents, des plateformes de visualisation et enfin des outils d'analyse. Ces travaux se font dans un univers extrêmement concurrentiel, avec une très large représentation de produits *open-source*. *A contrario*, solidement campée sur leur rente réglementaire, les éditeurs historiques de manuels scolaires ont très largement ignoré les avancées de la recherche en sciences cognitives et didactiques, se limitant à une simple numérisation des contenus sans valeur ajoutée supplémentaire (*adaptive learning, assessment, fouille de données*). Dans ce paysage éditorial, Canopé et le CNED sont particulièrement exposés à la concurrence de plateformes d'*e-Learning*. Certains de ces produits se prévalent déjà d'une estampille *Learning Analytics*, quand ils ne proposent encore que des outils élémentaires de *tracking* et de statistiques.

Face aux acteurs traditionnels des industries éducatives, avancent quatre multinationales massivement investies dans les technologies et techniques de fouille de données : Microsoft, IBM, Google et Amazon. Elles ont construit leur leadership sur leur capacité à adapter leurs algorithmes aux besoins du grand public (la recherche en ligne) comme aux problématiques métier de la finance et du commerce. Les *Analytics* dans l'éducation n'ont pas encore eu l'impact qu'elles ont dans d'autres domaines. Au moment où ces entreprises achèvent une transformation essentielle de leur modèle commercial, du logiciel au service, elles trouvent un nouveau marché dans l'Education. En France, des regroupements d'entreprises spécialisés dans la formation en ligne ont l'ambition de faire avancer les problématiques des *Learning Analytics* comme Educa Lab et la French Tech.

Face à ces acteurs commerciaux, il y a les institutions de la recherche et de l'éducation. Les premières sont très activement investies dans des programmes transdisciplinaires en sciences de l'éducation, en sciences cognitives et en informatique pour l'apprentissage humain. Promouvoir la généralisation des *Learning Analytics* exige de renforcer l'étroite collaboration entre les entreprises, la recherche et les institutions autour de projets ouverts. Prenons l'exemple de trois initiatives récentes :

*The LACE (Learning Analytics Community Exchange) Project* est porté par 9 universités européennes et financé par l'Union au titre du Septième Programme-Cadre de 2014 à juin 2016 [<http://www.laceproject.eu/>]. Le site web réunit des contributions sur l'usage des *Learning Analytics* dans le système secondaire de différents pays européens. Ce projet vise à réunir les communautés scientifiques, les acteurs commerciaux et *open-source* des *Learning Analytics* pour dégager ensemble des bonnes pratiques et atteindre quatre objectifs :

- promouvoir les échanges, l'innovation ;
- élargir les corpus de données ;
- dresser de nouvelles perspectives ;
- établir un consensus autour de l'interopérabilité et de l'échange de données.

La fondation *Apereo* est une émanation de la communauté *Sakai*, un ensemble d'institutions académiques, d'organisations commerciales et d'individus contribuant au développement de la plateforme d'apprentissage *open source Sakai CLE*. Parmi ses projets, *The Apereo Learning Analytics Initiative (LAI)* vise à

1. accélérer la réalisation de logiciels et de frameworks de *Learning Analytics* ;
2. soutenir la validation de pilotes dans les institutions partenaires ;



### 3. mutualiser autant que se peut les développements.

Mentionnons enfin *The Experience API* (ou *xAPI*, anciennement *Tin Can API*) : un projet, piloté par *Rustici Software* (US-TN), sous l'égide du Département de la Défense des Etats-Unis. *xAPI* cherche à promouvoir des standards internationaux permettant l'interopérabilité des données, d'un système à un autre [<https://experienceapi.com/>]. Cette interface de programme d'application (API, *Application program interface*) permet de collecter des données concernant les activités d'une personne ou d'un groupe sous la forme de triplets RDF (acteur, action, résultat), au sein comme en dehors d'un LMS ou d'un navigateur. Au-delà de l'apprentissage formel, *xAPI* permet de tracer l'apprentissage informel, social et les expériences du monde réel, comme lire un article, regarder une vidéo ou avoir une discussion avec un tuteur. Les triplets sont conservés dans un *Learning Record Store* (LRS comme *Learning Locker*), auquel accèdent des services tiers pour générer des analyses et des visualisations, à l'attention des différents acteurs de la chaîne éducative, de l'apprenant au gestionnaire.

Ces projets montrent la capacité des différents acteurs du marché éducatif à évoluer autour d'objectifs partagés. Les institutions (Etats, Commission européenne, Universités, Centres de recherche) conservent un rôle essentiel pour impulser et encadrer ces expérimentations, dans un environnement sensible, soumis aux multiples pressions des lobbyistes de l'industrie du logiciel.

### 3.2 *L'expérimentation: entre éthique, droit et évolution*

Jusque récemment, la collecte des données reste fragmentée, limitée et sommaire. Fragmentée car les données d'apprentissage restent localisées sur une multitude de LMS non reliés ; limitée car une faible proportion de cours sont dispensés via de tels dispositifs ; sommaires car les seules données conservées dérivent, dans la plupart des cas, du modèle SCORM (*Sharable Content Object Reference Model*), créé pour décrire des ressources et non des activités. Ainsi les enseignants et gestionnaires de ces plateformes de cours en ligne avaient-ils un accès sommaire aux taux de présence, de complétion, de réussite et de participation. Le cadre juridique a fait l'objet d'arrangements successifs avec la CNIL. La donnée a changé de statut, désormais disponible en temps réel et dans des volumes dont la croissance suit la loi de Moore. Appliqués au grand public, les techniques, méthodes et algorithmes venus de la Défense et du Marketing s'étendent à l'Education, agitant l'épouvantail d'une société de surveillance, d'un panoptique foucauldien au sein duquel les mineurs eux-mêmes, se sachant contrôlés, apprendraient l'autocensure<sup>2</sup>. Quelles données sont collectées ? Quel degré de confiance peut-on accorder à une solution logicielle lorsque les algorithmes sont protégés par le sceau des brevets industriels ?

Si la mesure ponctuelle du pouls d'un élève de cycle 3 est nécessaire à l'apprentissage d'une gestion correcte de l'accélération de son rythme cardiaque, est-il pour autant pertinent de l'équiper en permanence d'une montre connectée ? Est-il opportun d'évaluer les établissements scolaires sur la base de l'activité de leurs visiteurs géolocalisés sur les réseaux sociaux ? Peut-on discriminer une élève pour avoir refusé d'être tracée dans ses déplacements *via* une puce RFID (*Radio Frequency IDentification*) [THE RUTHERFORD INSTITUTE 2013]. A partir de quel pourcentage d'erreur est-on prêt à s'en remettre à la force algorithmique pour détecter les décrocheurs ? Les scientifiques doivent contribuer de deux façons à ce débat citoyen : en analysant les pratiques de terrain et en proposant des indicateurs ouverts. Le projet HUBBLE (*HUman oBservatory Based on anaLysis of e-Learning traces*) est une initiative financée par l'Agence nationale de la recherche [<http://hubblelearn.imag.fr/>]. Il propose la création d'un observatoire pour la construction et le partage de traces massives d'*e-learning*, de leurs processus d'analyse et de leurs contextes d'usage. Il s'agit d'un projet fédérateur et structurant autour des environnements d'*e-learning* massif (dont les MOOCS) comportant une dimension évaluative longitudinale pour comprendre les processus éducatifs afférents. Les processus d'analyse construits accompagneront la prise de décisions des acteurs intervenant dans le système d'enseignement et d'apprentissage ; ils devront être utilisables et redéployées dans leurs environnements ; ils guideront le chercheur en *e-learning* pour la production de concepts, de modèles et d'indicateurs.

De la maternelle au lycée, le public des apprenants est mineur et ne prend conscience que très progressivement des enjeux associés aux droits de la personne et au consentement. C'est la raison pour laquelle les responsables légaux doivent être sollicités. Les données relatives à la santé, aux capacités cognitives et à la domiciliation d'un mineur n'en restent pas moins protégées et considérées sensibles par la puissance publique. Des législations généralistes ont été adoptées, le FERPA (*Family Educational Rights and Privacy Act*) en 1974 aux Etats-Unis, la loi "Informatique et libertés" en 1978 en France, le *Data Protection Act* en 1998 au Royaume-Uni. Ces législations sont dans la ligne de mire des acteurs de l'économie numérique. Dans une perspective néo-libérale

---

<sup>2</sup> Dans *Surveiller et punir* (1975), Michel Foucault consacre tout un chapitre au panoptique, invention de Jeremy Bentham, philosophe et réformateur britannique. Le panoptique est une tour centrale dans laquelle se trouve un surveillant ; autour de cette tour des cellules sont disposées en cercle. Le panoptique, c'est finalement faire de la visibilité la prison. L'essentiel, c'est que l'on se sache surveillé. A l'inverse, le surveillant étant invisible, le prisonnier ignore s'il est surveillé ou non. Foucault étend ce principe aux ateliers de fabrication, aux pensionnats, aux casernes, etc.

d'*empowerment*, un intense lobbying plaide pour une émancipation des consommateurs. Le Forum Economique Mondial de Davos en 2013 publie un rapport en ce sens [WORLD ECONOMIC FORUM 2013]. Avec la promesse d'une croissance, le rapport appelle à libérer les données personnelles du carcan de législations contraignantes. Chose faite en France avec l'adoption de la loi Lemaire en janvier 2015 : c'est la personne fichée elle-même et non le législateur ou la CNIL qui décidera désormais de l'usage qui sera fait de ses données personnelles. Ce droit à l'autodétermination informationnelle présume la nécessaire maîtrise par l'individu de ses données. Dans le même temps, avec l'adoption de la Loi relative au renseignement (promulguée le 24 juillet 2015), l'Etat s'arrogue un pouvoir de surveillance soustrait au contrôle de la CNIL.

Nous en arrivons à une situation paradoxale. La collecte d'une adresse mail d'un mineur sur un site éducatif n'est pas toujours justifiable pour la CNIL (Commission Nationale de l'Informatique et des Libertés) dont la dernière étude de septembre 2015 montre clairement que les règles de la protection de la vie privée des enfants et adolescents sont largement ignorées par les éditeurs. Mais que peuvent valoir ces mesures de prudence lorsque par l'emploi d'un système d'exploitation, la navigation connectée ou l'installation d'une application, un apprenant mineur se dépossède du contrôle des données personnelles qui le concernent ? Est-ce le rôle des pouvoirs publics d'attacher les enfants à ces fournisseurs de service ? A-t-on prévu d'informer un élève, et ses parents, de sa capacité à se soustraire à toute expérimentation en ce sens ? Nous ne sommes jamais à l'abri d'une utilisation biaisée ou mal intentionnée. A Philadelphie, un système d'activation à distance de la webcam des macbooks à usage éducatif a été détourné par la direction d'un lycée pour stipuler à un élève que son comportement à la maison était inapproprié [CNN, 22 février 2010].

Ainsi, l'utilisation des données d'apprentissage des personnes mineures est devenue un véritable enjeu de politique générale. Aux Etats-Unis, les principaux industriels du marché ont souscrit un code de conduite censé rassurer le grand public : le *Student Privacy Pledge*. En Europe, la CNIL s'efforce de mener, avec ses homologues européens réunis en G29 (*Article 29 Working Party*), des activités de contrôle vis-à-vis des Etats et des multinationales. Toute institution s'engageant dans un projet de *Learning Analytics* devrait prendre en compte quatre axes de pilotage [JISC 2015] :

- Le premier axe consiste à définir des niveaux de responsabilité dans l'accès aux données et leur traitement. Figurent au sein du Comité d'éthique, un responsable général sur le plan juridique et éthique, et autant de responsables que de missions spécifiques : collecter les données, les anonymiser, appliquer des traitements, conserver et gérer les données personnelles. Ce comité devrait compter des représentants des élèves et des personnels clés des institutions concernées.
- De l'apprenant et ses parents, au personnel éducatif de l'établissement, aux gestionnaires et/ou administrateurs du rectorat, chaque acteur dispose d'un accès différencié aux données, du niveau personnel au niveau d'un établissement ou d'un district. Il ne suffit pas que ces données soient agrégées pour qu'elles puissent passer pour anonymisées.
- Le système informatique doit assurer une étanchéité horizontale, entre individus, et verticale, entre acteurs de différents niveaux : ils ne doivent pouvoir être identifiés ni à partir de métadonnées, ni par recoupements.

- Les *Learning Analytics* visent à favoriser la compréhension et l'amélioration des apprentissages : elles ne sauraient être détournées dans une visée de contrôle non consenti.

La philosophie des *Learning Analytics* n'est pas d'instaurer une société de surveillance dès les bancs de l'école. Les *Learning Analytics* doivent être utilisés dans une perspective formative, généralement de façon agrégée et anonymisée. Mais n'est-ce pas là une contradiction avec la personnalisation de l'aide ? La problématique de la sécurité et de la confidentialité n'est-elle semblable au domaine médical ? Les données sensibles relatives à la santé, aux capacités cognitives, à l'environnement familial doivent faire l'objet de précautions particulièrement drastiques en terme de niveau d'accès et de sécurité. Le recueil et la conservation de ces données doivent avoir été justifiés *a priori*. La collecte doit être effectuée avec sobriété dans un protocole de recherche dûment explicité dans un contrat de transparence, selon des méthodes publiées. Ainsi, l'ensemble de ces données doit servir à développer des interventions positives auprès des différents types d'acteurs et ne jamais être accessibles à de tierces parties. Parce que ces expérimentations sont souvent menées en partenariat avec des entreprises non-Européennes, le commanditaire devrait exiger une clause de non-transfert des données personnelles hors Union (*Safe Harbor*), en conformité avec la décision de la Cour de Justice de l'Union Européenne du 6 octobre 2015.

Enfin, dans l'approfondissement du socle des connaissances et des compétences, les apprenants exposés aux objets connectés développent une connaissance de leur fonctionnement. Sur la toile de fond d'une éducation au numérique, ils sont informés des enjeux de la neutralité de l'internet, de l'*open-source*, de l'aliénation du contrôle sur les données personnelles et de la force des algorithmes dans nos existences virtuelles désormais filtrées. Les personnels d'éducation et les enseignants, les parents et citoyens, doivent également être formés aux enjeux des *Learning Analytics*. Les *Learning Analytics* proposent une aide qui ne saurait, en aucun cas, se substituer à la réflexion et l'engagement humaniste de l'enseignant.

### 3.3 Les Learning Analytics en 2025 : huit scénarios d'évolution

Dai Griffiths a coordonné pour le réseau LACE (*Learning Analytics Community Exchange*) une consultation sur l'évolution des *Learning Analytics* à l'horizon 2025 [GRIFFITHS 2015]. Pour mener cette grande consultation les protagonistes ont mobilisé la méthode Delphi. Dans un premier temps ils ont développé des scénarios. Ensuite, le questionnaire a été soumis à des experts du secondaire, de l'université et du monde du travail sur trois principaux thèmes : technologie, confidentialité et éthique, pédagogie. Dans un troisième temps, les réponses ont été soumises aux décideurs pour analyser ce qui est de l'ordre du désirable et du faisable. Enfin, une analyse stratégique permet d'éclaircir les divergences entre experts et décideurs et identifier les fossés entre infrastructures et pratiques.

103 experts distribués en Amérique, Europe et Australasie ont été conduits à évaluer au moins 3 de ces scénarios via une échelle de Likert ainsi qu'à justifier leur évaluation par un commentaire libre (Figure 5). En mêlant adroitement considérations éthiques, évolution technologique et volonté politique, cette enquête montre que les experts ont davantage foi dans la capacité des technologies qu'ils ne s'accordent sur les objectifs à poursuivre (Tableau 8). Au terme de l'examen des réponses, cinq interrogations subsistent.

Un premier doute s'insinue sur la capacité des *Learning Analytics* à s'imposer dans le champ éducatif : si le scénario 3 est jugé très indésirable, de nombreux experts en envisagent la possibilité. Mais la vision 6 rassemble un double consensus sur l'échelle du désirable et du faisable. Il en ressort que les décideurs devraient sensibiliser à, et soutenir, une recherche ouverte, ancrée sur des modèles pédagogiques humanistes.

Deuxièmement, un fort consensus se dégage, sur le plan éthique, autour de la nécessité de réguler le contrôle des données d'utilisateurs pour renforcer les droits de la personne concernée. Le *Scénario 4. Les individus contrôlent leurs propres données* est considéré comme très désirable et pour la majorité amplement faisable, entérinant là les influents codes de bonnes pratiques publiés par JISC et l'Open University UK [JISC 2015, The Open University, 2014]. Pour autant, ce scénario exige des puissances publiques, organismes de réglementation et gouvernements d'imposer un cadre législatif à la collecte et l'usage des données et d'étendre cette politique de confidentialité aux outils.

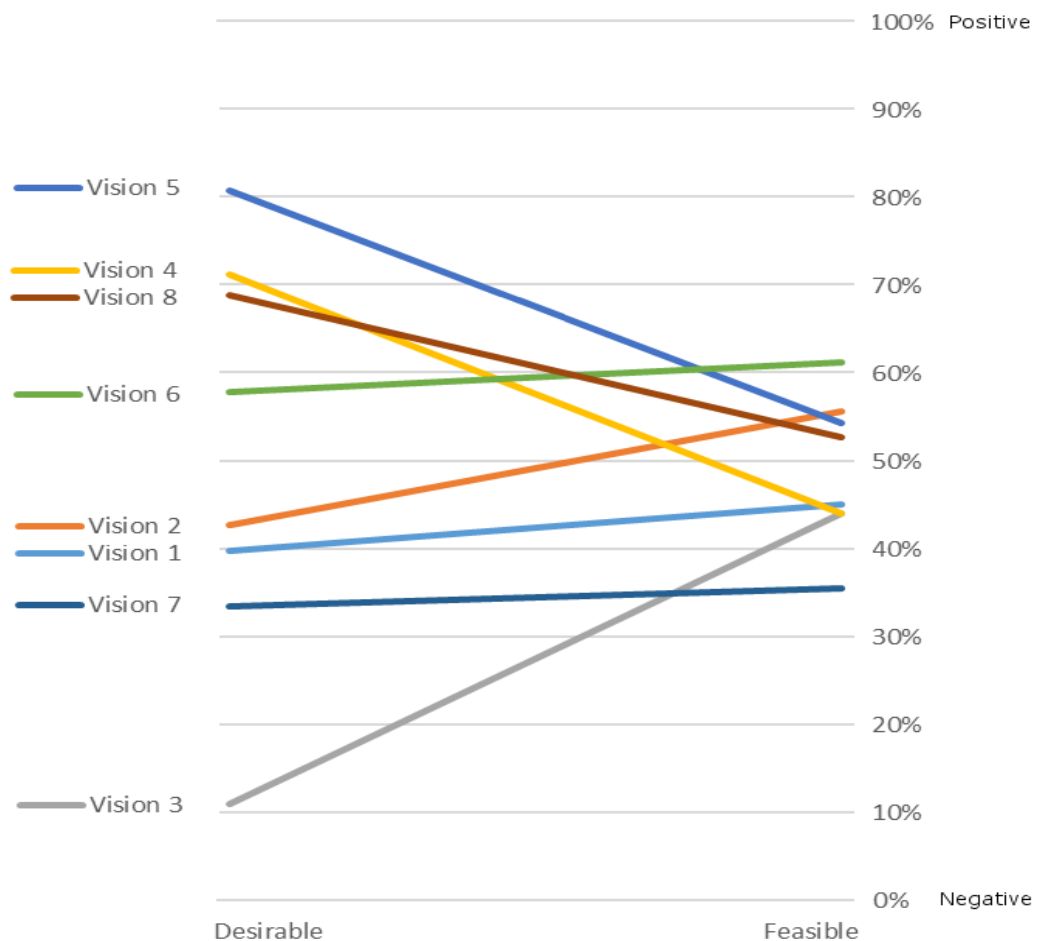
Troisièmement, la plus grande préoccupation des experts concerne le déploiement d'une pédagogie constructiviste, dirigée par un enseignant plutôt que focalisée sur les compétences et la maîtrise des contenus. Tandis que le doute persiste au sujet des indicateurs à utiliser, un consensus s'impose pour rappeler la place centrale de l'humain dans l'enseignement et dissocier *Learning Analytics* de l'automatisation de l'apprentissage. En ce sens le *Scénario 7. Les Analytics soutiennent un apprentissage autonome auto-dirigé* a fait l'unanimité contre lui, à l'inverse du Scénario 8. Cette préoccupation engage les décideurs à financer la recherche sur les usages pédagogiques des *Learning Analytics*, œuvrer au rapprochement des professionnels de l'éducation avec les informaticiens.

Quatrièmement, les experts disent leurs inquiétudes croisées au sujet des données : les mots 'pouvoir', 'confidentialité', 'législation' et 'éthique' pèsent pour plus d'un tiers des termes utilisés dans les commentaires libres : les bénéfices des Analytics pourraient être menacés par une réaction contre la collecte intrusive des données et l'usage inapproprié des résultats de l'analyse. Il convient donc de construire un consensus social et politique autour de la collecte et l'usage des données : développer des systèmes permettant aux personnes concernées de

contrôler leurs données, promouvoir la transparence et le sens de la responsabilité dans la collecte et l'usage. Les organismes de contrôle devraient pouvoir contrôler et imposer la confidentialité des données aux outils.

Cinquièmement, des différences s'insinuent entre experts. Le *Scénario 1. Les classes surveillent l'environnement physique* est davantage apprécié par les enseignants du primaire/secondaire que ceux du supérieur ou des experts de la formation au travail. Pour les répondants de l'enseignement scolaire et de la formation au travail le *Scénario 5. Des systèmes ouverts de Learning Analytics sont largement adoptés* semble inatteignable : la perception des *Learning Analytics* à l'Ecole est un chantier considérable. Enfin le Scénario 6 semble moins désirable et atteignable aux répondants de l'Ecole qu'aux autres : ces différences confirment l'existence d'un fossé entre experts et praticiens. Les systèmes et méthodes proposées ne rencontrent pas une adhésion entière de répondants se définissant comme attirés par les *Learning Analytics*. Développer la recherche-action autour des *Learning Analytics* devient une priorité.

**Figure 5. Faisable et/ou désirable ? L'évaluation des scénarios**



**Tableau 8. Les Learning Analytics (LA) en 2025 : huit scénarios à l'étude**

<p><b>1. Les classes surveillent l'environnement physique pour soutenir l'apprentissage et l'enseignement.</b></p> <ul style="list-style-type: none"> <li>• Cahiers, crayons, tablettes,... auront des capteurs.</li> <li>• Caméras à reconnaissance faciale traceront l'apprenant.</li> <li>• Données utilisées pour suivre le progrès de l'apprenant.</li> <li>• Prise en charge de l'apprentissage d'un large éventail de compétences physiques.</li> <li>• Les enseignants seront alertés des signes d'ennui, confusion ou éloignement de la tâche attendue.</li> <li>• Les enseignants suivront les interactions sociales numériques et pourront identifier où ils devront favoriser la socialisation et le comportement collaboratif.</li> </ul>
<p><b>2. La traçabilité des données personnelles soutient l'apprentissage</b></p> <ul style="list-style-type: none"> <li>• Des capteurs sophistiqués livreront des données personnelles : posture, stress, attention, niveau de sucre...</li> <li>• Les personnes collectent leurs informations et les utilisent dans leurs programmes pour avoir des recommandations sur leurs apprentissages.</li> <li>• Les apprenants récupèrent les statistiques et les données associées aux facteurs de réussite des apprentissages dans un domaine particulier.</li> <li>• Des institutions d'enseignement connues vendent des programmes, à partir de ces informations, pour optimiser l'apprentissage en fonction des âges et des cours.</li> <li>• Des entreprises font la même chose pour certains domaines.</li> <li>• Certains étudiants partagent leur savoir faire.</li> <li>• La plupart des institutions d'apprentissage suivent les programmes d'auto-supervision.</li> </ul>
<p><b>3. Les Analytics sont rarement utilisés dans l'éducation</b></p> <ul style="list-style-type: none"> <li>• Les cours qui sont automatisés, utilisant les LA, sont méprisés.</li> <li>• Les apprenants ont compris qu'ils peuvent « jouer » avec le système.</li> <li>• Il y aura un grand vide sur la gestion des données sensibles et des usages abusifs.</li> <li>• L'usage des données d'apprentissage devra être approuvé par l'apprenant et par des nouveaux inspecteurs.</li> <li>• Un consensus a émergé dans la politique éducative: l'abandon des LA</li> </ul>
<p><b>4. Les individus contrôlent leurs propres données</b></p> <ul style="list-style-type: none"> <li>• Les apprenants contrôlent             <ul style="list-style-type: none"> <li>• le type et la quantité des données qu'ils souhaitent partager.</li> <li>• Avec qui ces données sont partagées (personnes, institutions, entreprises,...)</li> <li>• La durée de la disponibilité de ces données.</li> </ul> </li> <li>• Des outils pour permettre ces contrôles sont clairement énoncés et faciles d'usage</li> <li>• Les institutions sont engagées dans la prise de conscience et reconnaissent les problèmes</li> </ul>
<p><b>5. Des systèmes ouverts de Learning Analytics sont largement adoptés</b></p> <ul style="list-style-type: none"> <li>• Les institutions éducatives demandent le contrôle des outils des LA             <ul style="list-style-type: none"> <li>• Comment les outils fonctionnent-ils ?</li> <li>• Pourquoi sont-ils utilisés ?</li> <li>• Elles conçoivent avec des fournisseurs des outils accompagnant leurs visions stratégiques.</li> </ul> </li> <li>• Il existe plusieurs fournisseurs ; tous utilisent des algorithmes ouverts et partagent les données suivant des standards qui facilitent la transparence et une validation indépendante.</li> <li>• Des visualisations, largement testées et accessibles, sont utilisées.</li> <li>• Les apprenants et enseignants peuvent utiliser en toute confiance une gamme d'outils variée.</li> </ul>
<p><b>6. En 2025, les systèmes de Learning Analytics sont des outils essentiels pour la gestion éducative</b></p> <ul style="list-style-type: none"> <li>• Les systèmes de recommandation des ressources sont largement utilisés.</li> <li>• Un grand nombre de données d'apprenant est utilisé pour générer des prédictions, de qualité et en temps réel, sur la probabilité de réussite.</li> <li>• Les apprenants et les enseignants planifient leur travail sur la base d'outils fiables qui recommandent et personnalisent ce qui doit être fait pour obtenir le meilleur résultat d'apprentissage.</li> <li>• Une industrie croissante offre des services institutionnels et individuels pour accompagner cette activité.</li> <li>• Précise, l'information prédictive permet aux gestionnaires et décideurs d'agrandir/réduire leur offre de formation</li> </ul>
<p><b>7. Les Analytics soutiennent un apprentissage autonome auto-dirigé</b></p> <ul style="list-style-type: none"> <li>• Développement d'un énorme corpus de données contenant l'information de millions d'apprenants.</li> <li>• Recommandations automatiques fiables, basées sur l'expérience de ces corpus, sur le meilleur chemin à prendre pour réussir l'apprentissage.</li> <li>• Les recommandations sont mieux informées et plus fiables que celles des meilleurs formateurs humains.</li> </ul>
<p><b>8. Les Analytics soutiennent un apprentissage autonome auto-dirigé</b></p> <ul style="list-style-type: none"> <li>• Disparition des curricula.</li> <li>• Les étudiants créent leurs groupes et décident leurs objectifs d'apprentissage et les moyens pour les atteindre.</li> <li>• Les Analytics accompagnent les échanges d'information et la collaboration entre les groupes.</li> <li>• Les enseignants deviennent des mentors.</li> <li>• Le suivi formatif est utilisé pour guider les progrès vers les objectifs fixés.</li> </ul>

## **Conclusion. L'Analytique des apprentissages numériques : une équation à trois inconnues.**

Les *Learning Analytics*, que nous proposons d'appeler analytique des apprentissages numériques, sont un terrain d'innovations fertiles : deux communautés, EDM et LAK, animent la recherche autour d'objectifs communs, d'approches complémentaires, et ce, en dépit d'une hybridation encore modeste. Leur impact scientifique est désormais conforté par des conférences annuelles, la publication de deux revues et, surtout, la création dans les plus prestigieuses institutions de l'enseignement supérieur et de la recherche de groupes de travail, comités de surveillance, centres de recherches, chargés de veiller à l'éthique, la déontologie, la finalité pédagogique de la collecte des données comme des algorithmes.

L'Analytique des apprentissages numériques s'attache à révéler à différents niveaux d'analyse les informations pertinentes pour améliorer l'expérience et les environnements d'apprentissage. Au service de projets humanistes, cet ensemble de technologies propose d'étayer les acteurs de la communauté éducative. Prédiction de la progression, analyse de l'apprentissage social, analyse de discours, tableaux de bord sont désormais des domaines d'expérimentation bien balisés. En tant qu'analytique des données résultant de l'apprentissage, l'analytique des apprentissages numériques s'apparentent au plus vaste ensemble de la Science des Données de l'Apprentissage, dont les questions de recherche vont de l'architecture physique des données aux traces issues de l'activité de l'enseignant lui-même.

Le passage de l'expérimentation à l'industrialisation est bousculé par l'arrivée imprévue des géants de l'économie numérique. En l'état actuel, la recherche expérimentale se teinte d'un certain scepticisme. Les garde-fous sont plus que jamais nécessaires s'agissant d'un public, pour partie, mineur et de données cognitives sensibles. Les équipes pluridisciplinaires de la recherche publique préconisent des dispositifs responsables, transparents, sobres et confidentiels, accessibles. Le futur de l'analytique des apprentissages numériques est une équation à trois inconnues : Premièrement, les décideurs (gouvernements, institutions, firmes) prendront-ils la mesure des financements et des projets encore nécessaires à l'adaptation de la recherche aux pratiques de terrain et à la dissémination d'une culture de l'analytique des données au sein du corps enseignant ? Ensuite, ces décideurs parviendront-ils à élaborer un cadre législatif viable, souple mais équitable, respectant le droit des usagers à suspendre la collecte des données et préservant leur confidentialité ? Enfin, ces décideurs parviendront-ils, au-delà d'une logique mercantile et de technologies propriétaires, à promouvoir des standards, des projets ouverts ?

Institutions d'enseignement et de recherche, politiques et chercheurs ont la main sur le devenir de l'analytique des données résultant de l'apprentissage numérique : de la connaissance de ce domaine et de la prise de conscience de ses impacts et enjeux sociétaux dépendra le visage de l'Ecole Numérique en 2025.



# Annexes

**Tableau 9. Présidence scientifique des conférence EDM (*Program Chairs*)**

Conférences	Présidence ( <i>Program chairs</i> ) & Institutions de rattachement	
EDM 16, Raleigh (US-NC), 29 juin – 2 juillet 2016 <a href="http://www.educationaldatamining.org/EDM2016/">http://www.educationaldatamining.org/EDM2016/</a>	Min Chi Mingyu Feng	North Carolina State University (US-NC) SRI International (US-CA)
EDM 15, Madrid (ES), 26-29 juin 2015 <a href="http://www.educationaldatamining.org/EDM2015/">http://www.educationaldatamining.org/EDM2015/</a>	Cristóbal Romero Mykola Pechenizkiy	University of Cordoba (ES) Eindhoven University of Technology (NL)
EDM 14, London (UK), 4 au 7 juillet 2014 <a href="http://educationaldatamining.org/EDM2014/">http://educationaldatamining.org/EDM2014/</a>	John Stamper - Zachary Pardos -	HCI Institute, Carnegie Mellon (US-PA) University of California Berkely (US-CA)
EDM 13, Memphis (TN), du 6 au 9 juillet 2013 <a href="https://sites.google.com/a/iis.memphis.edu/edm-2013-conference/">https://sites.google.com/a/iis.memphis.edu/edm-2013-conference/</a>	Sidney D’Mello Rafael A. Calvo	University of Notre Dame (US-IN) The University of Sydney (AU)
EDM 12, Chania, (GR), 19-21 juin 2012 <a href="http://educationaldatamining.org/EDM2012/">http://educationaldatamining.org/EDM2012/</a>	Kalina Yacef Osmar Zaiane	The University of Sydney (AU) University of Alberta (CA)
EDM 11, Eindhoven (NL), 6-8 juillet 2011 <a href="http://educationaldatamining.org/EDM2011/">http://educationaldatamining.org/EDM2011/</a>	Cristina Conati Sebastian Ventura	University of British Columbia (CA) University of Cordoba (ES)
EDM 10, Pittsburgh (US-PA), 11-13 juin 2010	Agathe Merceron, Philip I. Pavlik Jr.	Beuth University of Applied Sciences (DE) Carnegie Mellon University (US-PA)
EDM09, Cordoue (ES), 1 <sup>er</sup> -3 juillet 2009 <a href="http://www.educationaldatamining.org/EDM2009/">http://www.educationaldatamining.org/EDM2009/</a>	Tiffany Barnes Michel Desmarais	Univ. of North Carolina Charlotte (US-NC) Ecole Polytechnique de Montreal (CA)
EDM08, Montreal (CA), 20-21 juin 2008 <a href="http://www.educationaldatamining.org/EDM2008/">http://www.educationaldatamining.org/EDM2008/</a>	Ryan S. J. de Baker Joseph E. Beck,	Carnegie Mellon University (US-PA) Worcester Polytechnic Institute (US-MA)

**Tableau 10. Présidence scientifique des Conférences LAK**

Conférences	Présidence ( <i>Program chairs</i> ) & Institutions de rattachement	
LAK16, Edinburgh (UK) 25-29 avril 2016 <a href="http://lak16.solaresearch.org/">http://lak16.solaresearch.org/</a>	Shane Dawson Hendrik Drachler Car. Penstein Rosé	University of South Australia (AU) Open University of the Netherlands (NL) Carnegie Mellon University (US-PA)
LAK15, Poughkeepsie (US-NY) 16-20 mars 2015 <a href="http://lak15.solaresearch.org/">http://lak15.solaresearch.org/</a>	Paulo Blikstein Agathe Merceron George Siemens	Stanford University, (US-CA) Beuth Univ. of Applied Sciences (DE) University of Texas, Arlington, (US-TX)
LAK14, Indianapolis (US-IN), 24-28 mars <a href="https://lak14indy.wordpress.com/">https://lak14indy.wordpress.com/</a>	Stephanie Teasley Abelardo Pardo	University of Michigan (US-MI) The University of Sydney (AU)
LAK13, Louvain (BE), 8-12 avril 2013 <a href="https://lakconference2013.wordpress.com/">https://lakconference2013.wordpress.com/</a>	Dan Suthers Katrien Verbert	University of Hawai’i, US University of Leuven, BE
LAK12, Vancouver (CA-BC), 29.4-2 mai <a href="http://lak12.sites.olt.ubc.ca/">http://lak12.sites.olt.ubc.ca/</a>	Dragan Gasevic Simon Buckingham Shum	Athabasca University (CA) Open University (UK)
LAK11, Banff (CA-AB), 27 fev.-1 <sup>er</sup> mars <a href="https://tekri.athabascau.ca/analytics/">https://tekri.athabascau.ca/analytics/</a>	Gráinne Conole Dragan Gašević	Open University (UK) Athabasca University (CA)

**Tableau 11. Chercheurs par pays dans les *Program Committees* d'EDM et LAK (2009-2016)**

RANG	PAYS	TOTAL	EDM	LAK
1	USA	321	230	91
2	Canada	118	46	72
3	Australia	88	34	54
4	Spain	81	57	24
5	Germany	61	18	43
6	UK	58	24	34
7	Netherlands	44	18	26
8	France	39	22	17
9	Austria	19	10	9
10	New Zealand	18	9	9
11	Switzerland	18	6	12
12	Belgium	18	5	13
13	Greece	13	2	11
14	Romania	11	6	5
15	Serbia	9	2	7

**Tableau 12. Top 10 des universités européennes en nombre de *reviewers* aux conférences EDM et LAK**

Nombre de relectures	Universités	Reviewers
43	University of Córdoba	Cano A., García-Martínez C., Gibaja E., Luna J., Luque M., Morales C., Olmo J., Romero C., Romero J., Soto S., Ventura S., Zafra-Gómez A.
22	Open University of the Netherlands	Drachsler H., Kirschner P., Laat M., Rosé C., Scheffel M., Specht M.
17	The Open University of the UK	Buckingham-Shum S., Clow D., Conole G., Dietze S., Ferguson R., Liddo A., McAndrew P., Weller M., Whitelock D., Wild F.
17	Eindhoven University of Technology	Calders T., Knutov E., Pechenizkiy M., Verbert K.
14	Katholieke Universiteit Leuven	Berendt B., Duval E., Klerkx J., Verbert K.
13	University of Maine, Le Mans	Choquet C., Iksal S.
11	University of Lugano	Mazza R.
10	Beuth University of Applied Sciences Berlin	Berendt B., Merceron A., Pinkwart N.
8	The Open University	Dietze S., Ferguson R., McAndrew P., Rienties B.
8	Université Joseph Fourier Grenoble puis Université Pierre et Marie Curie, Paris	Luengo V.

**Tableau 13. Auteurs publiés dans les deux revues JEDM & JLA, de 2009 à 2016**

Auteurs	JEDM	JLA
Baker R.	BAKER, YACEF 2009. GOBERT <i>et al.</i> 2012. WINNE, BAKER 2013	PARDOS <i>et al.</i> 2014. WANG <i>et al.</i> 2014. KOVANOVIC <i>et al.</i> 2015.
Biswas G., Kinnebrew J. S	KINNEBREW <i>et al.</i> 2013	YE, BISWAS 2014. SEGEDY <i>et al.</i> 2015
Blikstein P.	SCHNEIDER, BLIKSTEIN, 2015	WORSLEY, BLIKSTEIN 2014. MERCERON <i>et al.</i> 2015
Graesser A. C.	FORSYTH <i>et al.</i> 2013.	DOWELL, GRAESSER 2014
Schneider B.	SCHNEIDER, BLIKSTEIN, 2015	SCHNEIDER, PEA 2015
Recker M.	XU, RECKER 2011. XU <i>et al.</i> 2013.	YUAN, RECKER 2014
Winne Ph. H.	WINNE, BAKER 2013	ALI <i>et al.</i> 2014. ROLL, WINNE 2015.
Yacef K.	BAKER, YACEF 2009	MARTINEZ-MALDONADO <i>et al.</i> 2015

# Bibliographie

La bibliographie est présentée selon le format APA, recommandé par la revue STICEF (*Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*).

Abbréviations :

*JEDM* : *Journal of Educational Data Mining*.

*JLA* : *Journal of Learning Analytics*.

- Aguiar, E., Ambrose, G. A., Chawla, N. V., Goodrich, V. et Brockman, J. (2014). Engagement vs Performance: Using Electronic Portfolios to Predict First Semester Engineering Student Persistence. *JLA*, 1(3), 7-33.
- Ahn, J. (2013). What can we learn from Facebook activity? Using social learning analytics to observe new media literacy skills. Dans *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK'13), 8-13 April 2013, Leuven, Belgium* (p. 135-144), New-York, ACM.
- Aleven, V., McLaren, B., Roll, I. et Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International journal of artificial intelligence in education*, 16(2), 101-128.
- Ali L., Hatala, M., Gašević, D. et Winne, Ph. H. (2014). Leveraging MSLQ Data for Predicting Students Achievement Goal Orientations. *JLA* 1(3), 157-160.
- Amershi S. et Conati C. (2009). Combining Unsupervised and Supervised Machine Learning to Build User Models for Exploratory Learning Environments. *JEDM*, 1(1), 71-81.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA : Harvard University Press.
- Anderson, T. (1996). The virtual conference: Extending professional education in cyberspace. *International Journal of Educational Telecommunications*, 2(2/3), 121-135.
- Arnold, K. E et Pistilli, M. D. (2012). Course signals at Purdue: using learning analytics to increase student success . Dans *Proceedings of the 2nd international conference on learning analytics and knowledge* (p. 267-270), New York, NY : ACM.
- Arnold, K.E. (2010). Signals: Applying academic analytics. *Educause Quarterly*, 33, 1-10.
- Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., ... Woolf, B. P. (2007). Repairing disengagement with non-invasive interventions. Dans *Proceedings of the International Conference on Artificial Intelligence in Education* (p. 195–202), IOS Press.
- Baker R., Gowda, S. M. et Corbett, A. T. (2011). Automatically detecting a student's preparation for future learning: Help use is key. Dans *Proceedings of the 4th international conference on educational data mining* (p. 179–188), Eindhoven : Technische Universiteit Eindhoven.
- Baker, R., Corbett, A. T., Koedinger, K. R., Evenson, S. E., Roll, I., Wagner, A. Z., ... Beck, J. (2006). Adapting to when students game an intelligent tutoring system. Dans *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (p. 392–401), Springer-Verlag Berlin Heidelberg.
- Baker, R., de Carvalho, A. M. J. A., Raspat, J., Aleven, V., Corbett, A. T. et Koedinger, K. R. (2009). Educational Software Features that Encourage and Discourage "Gaming the System". Dans *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (p. 475- 482), IOS Press.
- Baker, R. et Siemens G. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. Dans *Proceedings of the 2nd international conference on learning analytics and knowledge* (p.252-254), New York , NY : ACM.
- Baker, R. et Siemens, G. (2014). Educational Data Mining and Learning Analytics. Dans K. Sawyer (dir.), *Cambridge Handbook of the Learning Sciences: 2<sup>nd</sup> Edition* (p. 253-274), New York , NY : Cambridge University Press .
- Baker, R. et Yacef K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM*, 1(1), 3-17.
- Balacheff, N. (1994). Didactique et intelligence artificielle. *Recherches en Didactique des Mathématiques, La Pensée Sauvage*, 14, 9-42.
- Balacheff, N. (1995). Conception, propriété du système sujet/milieu. Dans R. Noirfalise , M. J. Perrin-Glorian (dir.), *Actes de la VII<sup>e</sup> Ecole d'été de didactique des mathématiques* (p.215-229), Clermont-Ferrand : IREM de Clermont-Ferrand.

- Balacheff, N. et Gaudin, N. (2002). Students conceptions: an introduction to a formal characterization. *Cahier Leibniz* 65, 1-21.
- Balacheff, N. et Lund, K. (2013). Multidisciplinary vs Multivocality: The case of “Learning Analytics”. Dans *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK'13)*, 8-13 April 2013, Leuven, Belgium (p. 5-13), New-York, NY : ACM.
- Beal, C.R., Qu, L. et Lee, H. (2006). Classifying learner engagement through integration of multiple data sources. Dans *AAAI'06. Proceedings of the 21st National Conference on Artificial Intelligence* (p. 151-156), New York NY : ACM.
- Beal, C.R., Qu, L. et Lee, H. (2008). Mathematics motivation and achievement as predictors of high school students' guessing and help-seeking with instructional software. *Journal of Computer Assisted Learning*, 24, 507-514.
- Ben-Naim, D., Bain, M. et Marcus, N. (2009). A user-driven and data-driven approach for supporting teachers in reflection and adaptation of adaptive tutorials. Dans *Proceedings of the 2nd International Conference on Educational Data Mining* (p. 21-30).
- Benzekri, J. P. (1973). *Analyse des données*. Paris : Dunod.
- Bienkowski, M., Feng, M. et Means, B. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics : An Issue Brief*, Washington, D.C. : US Department of Education, Office of Educational Technology.
- Borgatti, S. P., Mehra, A., Brass, D. J. et Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892-895.
- Bowers, A. J. (2010). Analyzing the Longitudinal K-12 Grading Histories of Entire Cohorts of Students: Grades, Data Driven Decision Making, Dropping Out and Hierarchical Cluster Analysis. *Practical Assessment, Research & Evaluation*, 15(7), 1-18.
- Brown, M. (2012). *Learning Analytic : Moving from Concept to Practice*. Récupéré du site Educause Learning Initiative, collection ELI Briefs.
- Brynjolfsson, E., Lorin M. H. et Heekyung H. K. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? Récupéré du site *Social Science Research Network*, Working Paper Series, April 22, 2011.
- Buckingham Shum, S. et Ferguson, R. (2012). Social Learning Analytics. *Journal of Educational Technology & Society*, 15(3), 3-26.
- Calvo, R. A. et D'Mello, S. K. (2011). *New Perspectives on Affect and Learning Technologies*. New York, NY : Springer.
- Campbell J. P., DeBlois P. B. et Oblinger D. G. (2007). *Academic Analytics: A New Tool for a New Era*. Récupéré du site *EDUCAUSE Review*, 42(4), 40-57.
- Cen, H., Koedinger, K. et Junker, B. (2006). Learning Factors Analysis - A general method for cognitive model evaluation and improvement. Dans *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (p. 164-175), Springer-Verlag Berlin Heidelberg.
- Cen, H., Koedinger, K. et Junker, B. (2008). Comparing two IRT models for conjunctive skills. Dans *9th International Conference on Intelligent Tutoring Systems, ITS 2008, Montreal, Canada, June 23-27, 2008, Proceedings* (p. 796-798) : Springer-Verlag Berlin Heidelberg.
- Charleer, S., Klerkx, J. et Duval, E. (2014a), Learning dashboards. *JLA*, 1(3), 199-202.
- Charleer, S., Santos, J. L., Klerkx, J. et Duval E. (2014b). Improving teacher awareness through activity, badge and content visualizations. Dans Y. Cao, T. Våljataga, J. K. T. Tang, H. Leung and M. Laanpere (dir.), *New Horizons in Web Based Learning: Proceedings of the 1st International Workshop on Open Badges in Education* (p. 143-152), Springer International Publishing.
- Clauset, A., Newman, M. E. J. et Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), id 066111.
- Clow, D. (2013). MOOCs and the funnel of participation. Dans *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (p. 185-189), New York, NY : ACM.
- CNN (2010). *FBI investigates allegations webcam used to monitor student*. Article publié le 22 février 2010. URL : <http://edition.cnn.com/2010/CRIME/02/20/laptop.suit/?hpt=T2>
- Cooper, A. (2012). *A Brief History of Analytics*, The University of Bolton, CETIS Analytics Series 1(9). Récupéré sur <http://publications.cetis.org.uk/2012/529>

- Corbett, A.T. et Anderson, J.R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, 4, 253–278.
- D’Mello S., Olney A. et Person N. (2010). Mining Collaborative Patterns in Tutorial Dialogues. *JEDM*, 2, 2-37.
- D’Mello, S. K. et Graesser, A. C. (2012). Dynamics of Affective States during Complex Learning. *Learning and Instruction*, 22, 145-157.
- Davenport, Th. H., Haris, J. G. et Morison, R. (2010). *Analytics at Work. Smarter Decisions, Better Results*. Boston, MA : Harvard Business Press.
- Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Journal of Educational Technology & Society*, 11(3), 224-238.
- Dawson, S., Gašević, D., Siemens, G. et Joksimovic, S. (2014). Current state and future trends: A citation network analysis of the learning analytics field. Dans *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK’14)*, 24-28 March, Indianapolis, Indiana, USA (p. 231-240), New-York, NY : ACM.
- De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M. et Cannavacciuolo, L. (2011). Discourse-centric learning analytics. Dans *LAK 2011: 1st International Conference on Learning Analytics & Knowledge*, 27 Feb - 01 Mar 2011, Banff, Alberta (p. 23-33), New-York, NY : ACM.
- Dekker, G., Pechenizkiy, M. et Vleeshouwers, J. (2009). Predicting students drop out: a case study. Dans *Proceedings of 2nd international conference on educational data mining* (p 41–50).
- Dowell, N. M. M. et Graesser, A. C., (2014). Modeling Learners’ Cognitive, Affective, and Social Processes through Language and Discourse. *JLA* 1(3), 183-186.
- Duval, E. (2011). Attention Please! Learning Analytics for Visualization and Recommendation. Dans *LAK ’11 Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (p. 9-17), New-York, NY : ACM.
- Essa, A. et Ayad, H. (2012). Student success system: risk analytics and data visualization using ensembles of predictive models. Dans *LAK ’12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (p. 158-161), New-York, NY : ACM.
- Fayyad, U., Piatetsky, S. G. et Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37-54.
- Feng, M., Heffernan, N. et Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User modeling and user-adapted interaction*, 19(3), 243–266.
- Ferguson, R. (2009). *The construction of shared knowledge through asynchronous dialogue*. Thèse de doctorat, Milton Keynes : The Open University. URL : <http://oro.open.ac.uk/19908/>
- Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges*. Récupéré du site Educause Learning Initiative, URL : <http://kmi.open.ac.uk/publications/techreport/kmi-12-01>
- Ferguson, R. et Clow, D. (2015). Consistent Commitment: Patterns of Engagement across Time in Massive Open Online Courses (MOOCs). *JLA*, 2(3), 55-80.
- Ferguson, R., Clow, D., Beale, R., Cooper, A. J., Morris, N., Bayne, S. et Woodgate, A. (2015). Moving through MOOCs: pedagogy, learning design and patterns of engagement. Dans T. Klobucar, G. Conole *et alii* (dir.), *Design for Teaching and Learning in a Networked World* (p. 70–84), Springer.
- Ferguson, R., Cooper, A., Drachsler, H., Kismihók, G., Boyer, A., Tammets, K. et Martinez Monés, A. (2015). Learning Analytics: European Perspectives. Dans *5th International Learning Analytics & Knowledge Conference (LAK15)*, 16-20 March 2015, Poughkeepsie (p. 69-72). New York, NY : ACM.
- Forsyth, C. M., Graesser, A. C., Pavlik, Ph. Jr., Cai, Z., Butler, H., Halpern, D. et Millis, K. (2013). Operation ARIES!: Methods, Mystery, and Mixed Models: Discourse Features Predict Affect in a Serious Game. *JEDM*, 5(1) , 147-189.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75-174.
- Garrison, D. R., Anderson, T., et Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The internet and higher education*, 2(2), 87-105.
- Gasevic, D., Dawson S., Mirriahi N. et Long P. D. (2015). Learning Analytics – A Growing Field and Community Engagement. *JLA*, 2(1), 1-6.
- Gauthier, G., Frasson Cl. et VanLehn K. (dir.) (2000). *Intelligent Tutoring Systems. Proceedings of the 5th International Conference on Intelligent Tutoring Systems*. Berlin, Springer-Verlag.

- Gee, J. P. et Green, J. (1998). Discourse analysis, learning and social practice: a methodological study. *Review of Research in Education* 23, 119-169.
- Gobert, J. D., Sao Pedro, M. A., Baker, R., Toto, E. et Montalvo, O. (2012). Leveraging Educational Data Mining for Real-time Performance Assessment of Scientific Inquiry Skills within Microworlds. *JEDM*, 4(1), 111-143.
- Goldhaber, M. H. (1997). The Attention Economy and the Net. *First Monday*, 2(4).
- Griffiths D. (coord.) (2015). *Visions of the Future. Horizon Report*. Récupéré de [http://www.laceproject.eu/wp-content/uploads/2016/02/LACE\\_D3\\_2.pdf](http://www.laceproject.eu/wp-content/uploads/2016/02/LACE_D3_2.pdf)
- Hayashi, C. (1998). What is Data Science? Fundamental Concepts and a Heuristic Example. Dans Hayashi, C., Yajima, K., Bock, H.H., Ohsumi, N., Tanaka, Y., Baba, Y. (dir.), *Data Science, Classification, and Related Methods, Proceedings of the Fifth Conference of the International Federation of Classification Societies (IFCS-96)* (p. 40-51), Springer Japan.
- Haythornthwaite, C. et de Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. Dans *Proceedings of the 7th International Conference on Networked Learning* (p. 183-190), Aalborg, Denmark : Lancaster University.
- Hershkovitz, A, Nachmias, R. (2008). Developing a log-based motivation measuring tool. Dans *Proceedings of the First International Conference on Educational Data Mining* (p. 226-233).
- Jared, E. K. (2015). Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin. *JEDM*, 7(3), 18-67.
- Jisc (2015). *Code of practice for learning analytics*. Récupéré de <https://www.jisc.ac.uk/>. URL : [https://www.jisc.ac.uk/sites/default/files/jd0040\\_code\\_of\\_practice\\_for\\_learning\\_analytics\\_190515\\_v1.pdf](https://www.jisc.ac.uk/sites/default/files/jd0040_code_of_practice_for_learning_analytics_190515_v1.pdf)
- Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, Sh. et Graesser, A. (2015). How do you connect? Analysis of Social Capital Accumulation in connectivist MOOCs. Dans *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (p. 64-68), New-York, NY : ACM.
- Kay, J., Maisonneuve, N., Yacef, K. et Reimann, P. (2006). The Big Five and Visualisations of Team Work Activity. Dans *Proceedings of the International Conference on Intelligent Tutoring Systems* (p. 197-206), Heidelberg : Springer-Verlag Berlin.
- Kienle, A. et Wessner, M. (2006). The CSCL Community in its First Decade: Development, Continuity, Connectivity. *International Journal of Computer-Supported Collaborative Learning*, 1(1), 9- 33.
- Kinnebrew, J. S., Loretz, K. M. et Biswas, G. (2013). A Contextualized, Differential Sequence Mining Method to Derive Students' Learning Behavior Patterns. *JEDM*, 5(1), 190-219.
- Kizilcec, R. F., Piech, C., Schneider, E. (2013). Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. Dans *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (p. 170-179), New-York, NY : ACM.
- Knight, S., Littleton, K. (2015). Dialogue as Data in Learning Analytics for Productive Educational Dialogue. *JLA*, 2(3), 111-143.
- Knowles, J. E. (2015). Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin. *JEDM*, 7(3), 18-67.
- Koedinger, K. R., Corbett, A. (2006). Cognitives tutors: Technology bringing learning science to the classroom. Dans R. K. Sawyer (dir.), *The Cambridge handbook of the learning sciences* (p. 61-78), New York, NY : Cambridge University Press.
- Koedinger, K., McLaughlin, E. et Stamper J. (2012). Automated student model improvement. Dans *Proceedings of the 5th international conference on educational data mining* (p. 17-24).
- Koedinger, K., Pavlik, P., Stamper, J., Nixon, T. et Ritter, S. (2010). Avoiding problem selection thrashing with conjunctive knowledge tracing. Dans *Proceedings of the 3rd international conference on educational data mining* (p. 91-100).
- Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S. et Baker, R. (2015). Does Time-on-task Estimation Matter? Implications on Validity of Learning Analytics Findings. *JLA*, 2(3), 81-110.
- Lallé, S., Mostow, J., Luengo, V. et Guin, N. (2013). Comparing Student Models in Different Formalisms by Predicting their Impact on Help Success. *AIED 2013 - 16th International Conference on Artificial Intelligence in Education* (p. 161-170), Springer.
- Larusson, J. A. et White, B. (dir.) (2014). *Learning Analytics: From Research to practice*, New York : Springer.

- Levin, D. Z. et Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge. *Management Science*, 50(11), 1477-1490.
- Long, P. D. et Siemens G. (2011). Penetrating the Fog: Analytics in Learning and Education. *Educause Review*, 46(5), 30-40. Récupéré de <https://net.educause.edu/ir/library/pdf/ERM1151.pdf>
- Macfadyen, L. P. et Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2), 588-599.
- Martin, T., Aghababayan, A., Pfaffman, J., Olsen, J., Baker, S., Janisiewicz, P., ... Petrick Smith, C. (2013). Nanogenetic Learning Analytics: Illuminating Student Learning Pathways in an Online Fraction Game. *LAK'13, Proceedings of the Third International Conference on Learning Analytics and Knowledge* (p. 165-169), New York, NY : ACM.
- Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J. et Clayphan, A. (2015). LATUX: an Iterative Workflow for Designing, Validating and Deploying Learning Analytics Visualisations. *JLA*, 2(3), 9-39.
- Mazza, R. (2010). Visualization in Educational Environments. Dans C. Romero, S. Ventura, M. Pechenizkiy et R. Baker (dir.), *Handbook of Educational Data Mining*, (p. 9-26), CRC Press.
- Mercer, N. (2004). Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, 1(2), 137-168.
- Merceron, A., Blickstein, P. et Siemens, G. (2015). Learning Analytics: From Big Data to Meaningful Data. *JLA*, 2(3), 4-8.
- Ming Ming, C. et Nobuko, F. (2014). Statistical Discourse Analysis: A method for modeling online discussion processes. *JLA*, 1(3), 61-83.
- Ming, N. C. et Ming, V. L. (2012). Predicting Student Outcomes from Unstructured Data. *Proceedings of the 2nd International Workshop on Personalization Approaches in Learning Environments* (p. 11-16), CEUR Workshop Proceedings.
- Minh Chieu, V., Luengo, V., Vadcard, L. et Tonetti, J. (2010). Student modeling in complex domains: Exploiting symbiosis between temporal Bayesian networks and finegrained didactical analysis. *International Journal of Artificial Intelligence in Education*, 20, 269–301.
- Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., Levy, R. (2012). Design and Discovery in Educational Assessment: Evidence-Centered Design, Psychometrics, and Educational Data Mining, *JEDM*, vol. 4/1, pp. 11-48.
- Mitchell, J. et Costello, S (2000). *International e-VET market research report: A report on international market research for Australian VET online products and services*, Sydney, Australia: John Mitchell & Associates and Education Image.
- Mitrovic, A., Martin, B. et Suraweera, P. (2007). Intelligent Tutors for All: The Constraint-Based Approach. *IEEE Intelligent Systems*, 22, 38–45.
- Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10, 151-177.
- Naur, P. (1969). ‘Datalogy’, the science of data and data processes. Dans A. J. H. Morrell (dir.) *Information Processing 68, Proceedings of IFIP Congress 1968, Edinburgh, UK, 5-10 August 1968*, (vol. 2, *Hardware, Applications*, p.1383-1387), Amsterdam : North-Holland Pub. Co.
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), p. 8577–8696.
- Niemann, K., Wolpers, M., Stoitsis, G., Chinis, G. et Manouselis N. (2013). Aggregating social and usage datasets for learning analytics: data-oriented challenges. Dans *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (p. 245-249), New York, NY : ACM.
- Nugent R. et Nelson B. (2012). Evidence-centered Design for Diagnostic Assessment within Digital Learning Environments: Integrating Modern Psychometrics and Educational Data Mining. *JEDM*, 4(1), 1-10.
- Ochoa X., Suthers D., Verbert K. et Duval E. (2014). Analysis and Reflections on the Third Learning Analytics and Knowledge Conference (LAK 2013), *JLA*, 1(2), 5-22.
- Ohlsson, S. (1994). Constraint-based student modeling. *NATO ASI Series F Computer and Systems Sciences*, 125, 167–189.
- Ohlsson, S. (1996). Learning from performance errors. *Psychological review*, 103, 241.
- Pardos, Z. A., Baker R., San Pedro M., Gowda S. M. et Gowda S. M. (2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *JLA*, 1(1), 107-128.

- Pavlik, P.I., Cen, H. et Koedinger, K.R. (2009). Performance Factors Analysis—A New Alternative to Knowledge Tracing. *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (p. 531–538), Amsterdam : IOS Press.
- Pelánek, R. (2015). Metrics for Evaluation of Student Models. *JEDM*, 7(2), 1-19.
- Rebolledo-Mendez, G., Du Boulay, B., Luckin, R. et Benitez-Guerrero, E. I. (2013). Mining Data From Interactions With a Motivational-aware Tutoring System Using Data Visualization. *JEDM*, 5(1), 72-103.
- Reffay, C. et Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. *Computer Supported Collaborative Learning*, Jun 2003, Bergen, Norway (p. 343-352), Kluwer Academic Publishers.
- Roberge, A. (2013). *Le LMS, un marché en croissance*. Récupéré de Thot Cursus, <http://cursus.edu/dossiers-articles/articles/20991/lms-marche-croissance/#.VqxzPrLhDIX>
- Roll, I. et Winne, Ph. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *JLA*, 2(1), 7-12.
- Romero, C., Ventura, S. (2007). Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, 33, 135-146.
- Romero, C., Ventura, S. , Pechenizkiy, M. et Baker R. (dir.) (2011). *Handbook of Educational Data Mining*. Boca Raton : CRC Press.
- Santally, A. (2006). Personalisation in web-based learning environments. *International Journal of Distance Education Technologies*, 4(4), 15-35.
- Santos, J. L., Govaerts, S., Verbert, K., et Duval, E. (2012). Goal-oriented visualizations of activity tracking: A case study with engineering students. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (p. 143-152), New York, NY : ACM.
- Santos, J. L., Govaerts, S., Verbert, K., et Duval, E. (2013). Addressing learner issues with StepUp! An evaluation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (p. 14-22), New York, NY : ACM.
- Sawyer, R. K. (dir.) (2014a). *The Cambridge handbook of the Learning Sciences*, New York, NY : Cambridge University Press.
- Sawyer, R. K. (2014b). The future of learning: Grounding educational innovation in the learning sciences , Dans *The Cambridge Handbook of Learning Sciences* (p. 726-746), New York, NY : Cambridge University Press.
- Schneider B. et Pea R. (2015). Does Seeing One Another's Gaze Affect Group Dialogue? A Computational Approach. *JLA*, 2(2), 107-133.
- Schneider, B., Abu-El-Haija, S., Reesman, J., et Pea, R. (2013). Toward collaboration sensing: applying network analysis techniques to collaborative eye-tracking data. Dans *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (p. 107-111), New York, NY : ACM.
- Schneider, B. et Blickstein P. (2015). Unraveling Students' Interaction Around a Tangible Interface using Multimodal Learning Analytics. *JEDM*, 7(3), 89-116.
- Slater, N. (2014). *Learning analytics: The current state of play in UK higher and further education*, Bristol : JISC. Récupéré de [http://repository.jisc.ac.uk/5657/1/Learning\\_analytics\\_report.pdf](http://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf)
- Segedy, J. R., Kinnebrew, J. S. et Biswas, G. (2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *JLA*, 2(1), 13-48.
- Siemens, G. et Gasevic, D. (dir.) (2012). *Journal of Educational Technology and Society*, 15(3), Special issue on Learning and Knowledge Analytics.
- Siemens, G., Gasevic, D., Haythornthwaite C., Dawson S., Buckingham Shum S., Ferguson R., ... Baker R. (2011). *Open Learning Analytics: an integrated & modularized platform Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*. Récupéré de <http://solaresearch.org/OpenLearningAnalytics.pdf>
- Skrypnik, O., Joksimović, S., Kovanović, V., Gašević, D. et Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. *The International Review of Research in Open and Distributed Learning*, 16(3), 188-217.
- Snow, E. L., Allen, L. K., Jacovina, M. E., Crossley, S. A, Perret, C. A. et McNamara, D. S. (2015). Keys to Detecting Writing Flexibility Over Time: Entropy and Natural Language Processing. *JLA*, 2(3) , 40-54.
- Sprague, R. (1980). A Framework for the Development of Decision Support Systems. *MIS Quarterly*, 4(4), 1-25.



- The Rutherford Institute (2013). *Victory Update: San Antonio Officials Re-Admit Expelled High School Student Andrea Hernandez to Magnet School After Ending RFID Tracking Program*. Récupéré de [https://www.rutherford.org/publications\\_resources/on\\_the\\_front\\_lines/victory\\_update\\_san\\_antonio\\_officials\\_re\\_admit\\_expelled\\_high\\_school\\_student](https://www.rutherford.org/publications_resources/on_the_front_lines/victory_update_san_antonio_officials_re_admit_expelled_high_school_student)
- Tukey, J. W. (1977). *Exploratory Data analysis*, Reading, MA : Addison Wesley.
- Van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: balancing between overview and overload. *JLA*, 2(2), 138-162.
- Van Leeuwen, A., Janssen, J., Erkens, G. et Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education*, 90, 80-94.
- Verbert, K., Duval, E., Klerkx, J. et Govaerts, S. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500–1509.
- Verbert, K., Govaerts, S., Duval, E., Santos, J., Van Assche, F., Parra, G. et Klerkx, J. (2014). Learning dashboards : An overview and future research opportunities. *Personal and Ubiquitous Computing Journal*, 18, 1499–1514.
- Walker, E. (2012). *Primer on K-20 Education Interoperability Standards*, Software & Information Industry Association. Récupéré de <https://www.siiia.net/LinkClick.aspx?fileticket=zuf9QNK3BZ4%3D>
- Wang, Y., Paquette, L. et Baker, R. (2014). A Longitudinal Study on Learner Career Advancement in MOOCs. *JLA* 1(3), 203-206.
- Waters, A., Studer, C. et Baraniuk, R. (2014). Collaboration-Type Identification in Educational Datasets. *JEDM*, 6, 28-52.
- Wells, G. et Claxton, G. (dir.) (2002). *Learning for Life in the 21st Century*. Oxford : Blackwell.
- Wertsch, J. (1991), *Voices on the Mind: A Socio-Cultural Approach to Mediated Action*, Cambridge, MA : Harvard University Press.
- Wertsch, J. V. (dir.) (1985). *Culture, Communication and Cognition: Vygotskian perspectives*. New-York, NY : Cambridge University Press.
- Winne Ph. H. et Baker R. (2013). The Potentials of Educational Data Mining for Researching Metacognition, Motivation and Self-Regulated Learning. *JEDM*, 5(1), 1-8.
- World Economic Forum (2013). *Unlocking the Value of Personal Data: From Collection to Usage*. Récupéré de [http://www3.weforum.org/docs/WEF\\_IT\\_UnlockingValuePersonalData\\_CollectionUsage\\_Report\\_2013.pdf](http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf)
- Worsley, M. et Blickstein, P. (2014). Analyzing Engineering Design through the Lens of Computation. *JLA*, 1(2), 151-186.
- Worsley, M. et Blikstein, P. (2013). Towards the development of multimodal action based assessment. Dans *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)* (p. 94-101), New York, NY : ACM.
- Wright, M. C., McKay, T., Hershock, C., Miller, K. et Tritz, J. (2014). Better Than Expected: Using Learning Analytics to Promote Student Success in Gateway Science. *Change. The Magazine of Higher Learning*, 46(1), 28-34.
- Xu, B. et Recker M. (2011). Understanding Teacher Users of a Digital Library Service: A Clustering Approach, *JEDM*, 3, 1-28.
- Xu, B., Recker M., Qi X., Flann N. et Ye L. (2013). Clustering Educational Digital Library Usage Data: A Comparison of Latent Class Analysis and K-Means Algorithms. *JEDM*, 5(2), 38-68.
- Ye, C., Biswas G. (2014). Early Prediction of Student Dropout and Performance in MOOCs using Higher Granularity Temporal Information. *JLA*, 1(3), 169-172.
- Yuan, M. et Recker M. (2014). Characterizing User Behaviors and Resulting Products in an Online Educational Community: A Comparison between Novices and Elders. *JLA*, 1(3), 150-153.
- Zhu, Y. et Xiong, Y. (2015). Towards Data Science. *Data Science Journal*, 14, 8.
- Zimmermann, J., Brodersen K. H., Heinemann, H. R. et Buhmann, J. M. (2015). A Model-Based Approach to Predicting Graduate-Level Performance Using Indicators of Undergraduate-Level Performance. *JEDM*, 7(3), 151-176.