

# Étude préliminaire des pratiques de navigation dans un environnement d'apprentissage informatique et mesure de leur efficacité pédagogique

Gilles **Lavigne**

Instituto de Investigación y Desarrollo Educativo  
Universidad Autónoma de Baja California, México  
[gilles@uabc.mx](mailto:gilles@uabc.mx)

Javier **Organista-Sandoval**

Instituto de Investigación y Desarrollo Educativo  
Universidad Autónoma de Baja California, México  
[javor@uabc.mx](mailto:javor@uabc.mx)

Lewis **McAnally-Salas**

Instituto de Investigación y Desarrollo Educativo  
Universidad Autónoma de Baja California, México  
[mcanally@uabc.mx](mailto:mcanally@uabc.mx)

*Recherche scientifique avec données empiriques*

## Résumé

Inscrite dans une expérience plus vaste centrée sur l'à-propos d'introduire des objets d'apprentissage dans la formation en statistique des étudiants de premier cycle, cette recherche vise plus précisément à étudier l'usage effectif des ressources accessibles dans un environnement informatisé d'apprentissage. L'analyse met en jeu le registre des *logs* fournis par la plateforme Moodle. Les résultats montrent que l'analyse de la navigation recèle une information à portée pédagogique intéressante, relative à la définition de stratégies d'apprentissage en ligne et à la mesure de l'efficacité pédagogique de ces dernières. Il semble exister, en effet, une relation entre l'intensité d'utilisation des ressources en ligne et la note obtenue à la fin de la formation.

## Mots clés

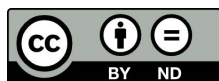
Trace, *log*, clic, apprentissage virtuel, statistique, Moodle, efficacité d'apprentissage

## Abstract

Carried out within an experiment aiming to evaluate the opportunity to introduce learning objects in undergraduate Statistics training, the present research investigated the relationship between students' use of learning resources in an online learning environment. Students' navigation patterns provide a rich source of pedagogical information about the learning process they enforced. Results show that the analysis of navigation patterns conceals information with learning range far more interesting, such as the definition of online training strategies and the measure of their learning efficiency. There seems to be a relation between the use intensity of the environment resources and the marks obtained by students.

## Keywords

Tracking, log, hits, virtual learning, Moodle, statistics, learning efficiency



©Auteur(s). Cette œuvre, disponible à [http://www.ritpu.org/IMG/pdf/RITPU\\_v05n01\\_06.pdf](http://www.ritpu.org/IMG/pdf/RITPU_v05n01_06.pdf), est mise à disposition selon les termes de la licence Creative Commons Attribution - Pas de Modification 2.5 Canada : <http://creativecommons.org/licences/by-nd/2.5/ca/deed.fr>

Cette étude s'inscrit dans le champ de la recherche sur l'apprentissage réalisé dans un environnement virtuel; elle explore les registres de navigation (logs) de la plateforme Moodle; elle vise à jauger la qualité informationnelle de ces données sur un plan pédagogique. L'adaptation de ces registres de navigation à des fins d'analyse est une opération complexe en raison du nombre élevé des données (dans ce cas Moodle fonctionne avec 154 bases de données), mais aussi de la diversité et de l'organisation mêmes des données; c'est pourquoi les résultats sont préliminaires et partiels.

## Introduction

L'étude utilise les registres de navigation d'étudiants ayant participé à une expérience *in situ* de la Universidad Autónoma de Baja California (UABC) au Mexique; celle-ci visait à évaluer l'opportunité d'utiliser des objets d'apprentissage dans des cours d'introduction à la statistique. Les étudiants engagés dans cette expérience suivaient leur formation selon une modalité mixte ou semi-présentielle (concept en vigueur au Mexique); les résultats de cette expérience ont été publiés.

Au cours d'une routine d'exploration de données, une relation quasi systématique, mais non significative sur le plan de la statistique, entre le nombre total de clics (*hits*) et les notes obtenues par les étudiants a motivé une première analyse dont les conclusions seront publiées sous peu (Organista, Lavigne et McAnally, sous presse). Ces résultats fragmentaires justifiaient une analyse plus poussée des caractéristiques de la navigation des étudiants et la recherche d'une possible assise pédagogique pour qualifier les regroupements répertoriés. C'est là l'objet de la présente étude, menée sur le plan méthodologique dans la mouvance du « forage de données » (*data mining*).

L'étude aborde la documentation par la formulation d'une problématique; elle décrit ensuite le protocole de recherche, dérivé de la méthodologie du forage de données, après avoir résumé une information contextuelle sur l'expérience génératrice des données; l'étude présente enfin les résultats et les discute, pour finalement en tirer des conclusions.

### 1. Formulation d'une problématique

Des divers courants de questionnement qui parcourent le champ de la recherche en éducation, plusieurs gravitent autour du problème de l'adaptabilité et de la flexibilité des environnements d'apprentissage informatiques. Un pôle d'intérêt : celui des usagers, professeurs, tuteurs, facilitateurs et étudiants, pour lesquels la recherche développe des indicateurs et des outils leur permettant d'intervenir dans le processus d'enseignement-apprentissage. La présente étude souscrit aux motivations de ce courant.

## 1.1 Les questions de recherche

Les registres (*logs*) consignent tous les mouvements de tous les étudiants dans l'espace informatique d'apprentissage d'un cours donné. Ils ont pour vocation première de fournir une information aux usagers, tuteurs, facilitateurs, professeurs, ainsi qu'aux étudiants eux-mêmes, sur la navigation de chacun d'entre eux, un à un, de jour en jour, de seconde en seconde, de page en page, de clic en clic. Cette information permet donc de suivre les parcours des apprenants et, selon la lecture qui en est faite, de détecter d'éventuels problèmes d'apprentissage individuels et d'intervenir dans le processus, ou encore de rétroagir sur le design pédagogique.

La présente étude examine trois aspects de la problématique de l'apprentissage en ligne : 1) Quelles variables caractérisent les pratiques de navigation des étudiants? 2) L'intensité d'utilisation des ressources Web par les étudiants permet-elle de les regrouper? 3) Les regroupements obtenus peuvent-ils s'interpréter selon une logique pédagogique? Comme les étudiants qui ont participé à l'expérience se trouvaient dans une situation d'apprentissage mixte, en présence et en ligne, se pose la question de l'impact de cette modalité.

## 1.2 La documentation scientifique

Les études utilisant les registres (*logs*) de plateformes informatiques d'apprentissage afin de documenter les pratiques de navigation des étudiants sont rares. Se pose, en plus, un problème linguistique d'adéquation terminologique. La documentation anglo-saxonne faisant état de recherches de ce type s'inscrit sous le couvert du « *educational data mining* »; la française sous celui de « traces d'apprentissage »; l'espagnole sous celui de « *minería de datos en educación* ». Le concept de « trace » semble peu utilisé en anglais et en espagnole, sinon en référence le plus souvent à des objets de recherche différents. Certes, un « log » peut être traité comme une « trace », mais une « trace » n'est pas nécessairement un « log »; le concept, d'ailleurs,

renvoie le plus souvent, soit à une « trace » explicite, comme un travail académique, soit au produit d'une observation, directe ou indirecte.

La documentation sur les traces est importante; dans son rapport de recherche bibliographique, Fischer (2005) identifie quelque 123 sources d'information vraisemblables traitant des traces d'apprentissage. Ce domaine de recherche sur les traces est de plus en plus structuré, dans l'Europe francophone tout au moins, grâce aux écoles thématiques du CNRS sur les EIAH, dont la 5<sup>e</sup> édition s'est tenue en 2007. De même, la documentation issue du *educational data mining* est aussi assez importante. Ce domaine se normalise, lui aussi, une première conférence internationale venant d'avoir lieu en juin 2008 à Montréal (de Baker, Barnes et Beck, 2008). La documentation en espagnol est plutôt succincte (Romero, Ventura et Hervás, 2005). Ni les unes ni les autres, toutefois, ne donnent accès à un savoir pertinent en relation avec cette étude.

## 1.3 Les traces et les logs comme données de recherche

Une partie de la recherche sur les traces vise à simplifier l'utilisation de l'information accessible en développant des indicateurs valides et efficaces (Cheype, 2006; Rozé, 2005; Settouti, 2006), en décelant des lacunes cognitives chez les apprenants ou des difficultés d'adaptation (Bisson, Bronner, Gordon, Nicaud et Renaudie, 2003; Vandebrouck et Cazes, 2004), en modélisant l'interprétation de ces traces dans le but de fournir à l'apprenant une information susceptible de l'aider dans ses apprentissages (Settouti, 2006) ou encore en offrant un support à la métacognition (Eyssautier-Bavay et Ollagnier-Beldame, 2006). Dans tous les cas, les résultats de ces recherches traitent de problématiques individuelles et s'adressent au tuteur, au professeur ou à l'étudiant.

Une des filières de recherche mettant en jeu des traces assimilables aux *logs* porte sur l'évaluation de l'utilisation des environnements informatiques par les étudiants (Buelens, Roosels, Wils et Van Rentergem, 2002). L'instrumentation développée étudie divers aspects de l'utilisation des ressources en ligne, de la navigation, de la consultation des rubriques, des travaux en ligne, de l'utilisation des hyperliens; il ressort à l'occasion que l'implantation d'environnements informatiques d'apprentissage ne donne pas toujours les résultats attendus (Buelens *et al.*, 2002). Ces recherches ont des visées administratives et non pédagogiques, et plusieurs auteurs soulignent le danger que de tels instruments deviennent de véritables logiciels espions (*spywares*) (Hardy, Bates, Antonioletti et Seed, 2005).

L'utilisation de *logs* comme données de recherche vise souvent à documenter, à renforcer ou à structurer le savoir sur un objet particulier de connaissance. Ainsi, le forage de données a été utilisé en relation avec les styles d'apprentissage (Durán et Costaguta, 2007) ou pour extraire une information pédagogique d'un tuteur informatique intelligent (Merceron et Yacef, 2005), pour analyser les comportements collectifs (Lancieri, 2007; Romero, Gutiérrez, Freire et Ventura, 2008), pour améliorer la scénarisation d'activités (Ferraris et Lejeune, 2007) ou encore pour prédire les réactions des étudiants (Feng et Heffernan, 2005, 2007). Mais une grande partie des recherches poursuit une visée technique d'expérimentation ou d'amélioration de procédures (de Baker *et al.*, 2008), de traitement des données (Merceron et Yacef, 2008), d'algorithmes (Romero, Ventura, Espejo et Hervás, 2008) ou simplement d'information (Bisson et Merceron, 2007). Là encore, l'apport documentaire en relation avec les questions de recherche reste faible.

#### 1.4 La modalité « mixte »

L'implantation vers la fin des années 1980 des nouvelles technologies de l'information et de la communication dans les universités a favorisé le développement de formations médiatisées et accessibles « en ligne » ainsi que la production d'instruments spécifiques pour ce faire. L'engouement pour cette nouvelle modalité d'enseignement-apprentissage n'a toutefois pas produit la vague innovatrice de changements annoncée au début des années 2000 (Zemsky et Massy, 2004).

L'introduction de cette nouvelle modalité pédagogique a, par contre, diversifié la recherche en éducation, en relation, entre autres, avec l'intégration des modalités « en présence » et « en ligne » dans les pratiques universitaires. L'élément clef de ce questionnement n'est autre que l'étudiant lui-même et sa perception de l'impact de cette modalité sur son apprentissage; celui-ci favorise l'interaction (Hardy *et al.*, 2005; Picciano, 2002), en particulier sociale (Lavigne, Organista et Aguirre Muñoz, 2006), ce qui renforce le bien-fondé de cette modalité à la fois mixte et collaborative (Backhoff, Lavigne, Organista et Aguirre Muñoz, 2007; Dziuban, Moskal et Hartman, 2004). Les définitions proposées requièrent encore une analyse théorique plus poussée (Charlier, Deschryver et Peraya (sous presse); Lavigne, Backhoff-Escudero et Organista, 2008). La recherche, toutefois, ne fournit aucune connaissance précise sur l'impact relatif des deux modalités.

#### 1.5 Les acquis documentés

Même si la documentation scientifique ne fournit pas de connaissances directement applicables dans l'étude proposée, c'est-à-dire relatives aux caractéristiques des types de navigation des étudiants dans un environnement d'apprentissage informatisé, elle offre néanmoins des balises importantes.

Ainsi est-il clair que l'analyse des *logs* des plateformes informatisées d'apprentissage constitue une voie pertinente de recherche à des fins de production de connaissances pédagogiques, tant sous forme d'indicateurs d'attitudes, de prédiction ou de styles d'apprentissage. Les études mettent en lumière, en effet, que le suivi et/ou l'observation de faits révélateurs des attitudes des étudiants dans un environnement informatique d'apprentissage constituaient une source d'information pédagogique importante pour les responsables des formations, tuteurs, professeurs, facilitateurs, et même pour les étudiants. Ce champ de recherche est en plein essor.

Par ailleurs, la méthodologie du forage de données montre que le recours aux processus d'échantillonnage, de segmentation, de classification et de mise en séquences des données était porteur d'une information pertinente et utile. Il ressort également, par contre, que l'emploi, *stricto sensu*, des techniques de la fouille échappe le plus souvent au personnel académique, dès lors client d'un service informatique. Ceci confirme, pour un, la nécessité de développer des outils, des modèles et des indicateurs, clés en main, mais aussi de banaliser, dans la mesure du possible, cette méthodologie.

L'impact relatif du présentiel et du virtuel dans une modalité d'enseignement-apprentissage mixte reste non quantifié; par contre, la pratique milite en faveur de gains concrets sur le plan de l'apprentissage réalisé dans le cadre de telles situations, mixtes ou hybrides. Néanmoins, dans le cadre de la présente étude, l'interprétation des influences réciproques des modalités reste en suspens.

## 2. Protocole de recherche

Les données analysées dans cette étude proviennent des registres (*logs*) de la plateforme Moodle, utilisée comme environnement d'apprentissage dans une expérience réalisée à la UABC lors du premier semestre de l'année 2006. Pour bien situer la qualité informationnelle de celles-ci, il est opportun de présenter les principales caractéristiques de cette expérience en tant que contexte de l'étude menée sur ces registres.

### 2.1 L'expérience *in situ* comme contexte

L'expérience donnait accès à trois leçons de statistique « en ligne » à des étudiants à l'intérieur du cours d'introduction qu'ils devaient suivre au cours du semestre. Ces trois leçons avaient été élaborées avec des objets d'apprentissage soit développés à cette fin, soit récupérés sur le Net, et elles étaient accessibles via la plateforme Moodle. La modalité d'enseignement-apprentissage était semi-présentielle, la composante virtuelle étant définie *a priori* comme support pédagogique aux étudiants et aux professeurs; cette modalité mixte était la seule possible puisque la UABC n'offre de cours ni totalement en ligne, ni totalement à distance, et que les syllabus en réglementent la diffusion. Tant les étudiants que les professeurs impliqués ont reçu une formation préalable d'initiation à l'usage de la plateforme. La participation était volontaire pour les étudiants. L'expérience visait à évaluer l'opportunité d'introduire des objets d'apprentissage dans la formation en statistique.

#### 2.1.1 Les participants

Les étudiants qui ont participé à l'expérience étaient inscrits à trois programmes de licence<sup>1</sup> constitutifs de trois des filières de la Faculté des sciences administratives et sociales (FACAYS, pour son sigle en espagnol). Le groupe I comptait 41 étudiants de sciences sociales, le groupe II, 24 d'administration et le groupe III, 27 d'informatique. Trois étudiants se sont joints en cours de route au troisième groupe pour un total de 95. Toutefois, plusieurs d'entre eux ne se sont pas présentés à la prise d'information prévue dans l'expérience; ceux du groupe III n'ont

étudié en ligne que la première leçon; la participation aux trois leçons n'a jamais été constante; il a été nécessaire d'échantillonner les étudiants pour certaines analyses; aussi le nombre total des participants aux diverses composantes de la recherche a-t-il varié. Aucun des étudiants n'avait déjà participé à une formation en ligne. Les enseignants responsables de chacun des trois cours de statistique ont agi comme instructeur de chaque groupe.

### 2.1.2 Les leçons

Les enseignants responsables des trois cours choisis pour l'expérience ont collaboré à la sélection, à la délimitation et à la composition des contenus des trois leçons : la courbe normale, la formulation d'hypothèse et les distributions du Z et du *t-student*. Une même structure de 10 rubriques organisait chacune des leçons : les rubriques 1, 2 et 3 pour les instructions, les rubriques 4 à 6 pour la matière, la rubrique 7 pour les outils, les rubriques 8 et 9 pour les exercices et la rubrique 10 pour le forum.

- 1) Objectifs
  - Objectifs pédagogiques de la leçon;
- 2) Route d'apprentissage
  - Proposition de cheminement d'apprentissage;
- 3) Contact
  - Adresses pour le soutien technique;
- 4) Concepts
  - Définitions des éléments constitutifs de l'objet de la leçon;
- 5) Cas résolus
  - Exemples de problèmes similaires résolus;
- 6) Information additionnelle
  - Hyperliens vers des libellés alternatifs de définition;
- 7) Outils
  - Programmes de simulation d'applications statistiques;
- 8) Problèmes
  - Formulation des problèmes à résoudre, avec hyperliens vers les autres rubriques;
- 9) Quiz-test
  - Exercices pour résoudre les problèmes, permettant plusieurs essais et comportant des rétroactions pédagogiques selon les erreurs commises;

### 10) Forum

- Non obligatoire pour les leçons 1 et 3, fortement recommandé pour la leçon 2.

À cette structure s'ajoutaient 3 rubriques intégrées hors leçon :

#### 1. Chat

- Clavardage;

#### 2. Répertoire

- Accès au répertoire des usagers et à leurs adresses de courriel;

#### 3. Évaluation

- Questions d'examen.

Lors de l'expérience, la page principale du cours donnait accès à toutes les rubriques des trois leçons et aux trois rubriques hors leçon, ainsi qu'aux services usuels, calendrier, événements, usagers en ligne... Aucun filtre ne conditionnait la navigation et les étudiants étaient libres de naviguer vers les destinations de leur choix. La rubrique « route d'apprentissage » donnait accès par un jeu d'hyperliens à toutes les rubriques d'une leçon.

### 2.1.3 Les sources de données

Les instruments de cueillette d'information, questionnaires pré- et post-expérience, et guides de discussion et d'entrevues, avaient pour objet de caractériser le profil socioéconomique des étudiants et de connaître leur trajectoire scolaire, ainsi que leur perception de l'apprentissage en ligne et de l'usage de la technologie, et leur attitude face à la statistique. Les réponses aux questionnaires, d'options multiples, étaient calibrées selon une échelle de Likert avec 5 pas entre le désaccord total et l'accord total. Les groupes de discussion réunissaient huit (8) étudiants chacun et les entrevues s'adressaient aux professeurs impliqués. Aux données recueillies avec ces instruments s'ajoutaient les relevés de notes et les archives de la plateforme Moodle.

### 2.1.4 La procédure

L'expérience a eu lieu durant la première session de 2006 dans les locaux de la FACAYS du campus universitaire de Valle Dorado de la UABC, dans la ville d'Ensenada (Mexique). Les trois cours de statistique, obligatoires, se sont déroulés de février à mai.

Les groupes I et II ont suivi les trois leçons, mais le groupe III n'a participé qu'à la leçon 1. Les professeurs responsables de chacun des trois groupes ont décidé du temps requis pour couvrir la matière de chaque leçon; tous révisaient les matières en classe; chacun d'eux a fait passer un examen partiel de connaissances au terme de chacune des trois leçons en ligne, examens qu'ils ont élaborés et notés selon leurs propres critères et coutumes.

## 2.2 La procédure d'analyse des logs

L'analyse des *logs* a été menée après la fin de l'expérience *in situ*. L'étude visait à explorer la navigation et l'usage que faisaient les étudiants de l'environnement informatique d'apprentissage, aussi le protocole de recherche ne s'est-il appuyé que partiellement sur la méthodologie de la fouille de données (Daedalus, 2008). La plateforme Moodle permet l'exportation de données vers Excel selon diverses modalités. Une fois importées, toutefois, ces données ont dû être nettoyées, pour certaines segmentées et pour d'autres échantillonnées; enfin, des analyses ont demandé le transfert de variables vers une base de données produite avec le programme d'analyse statistique *Statistical Package for the Social Sciences* (SPSS).

### 2.2.1 La population

L'étude, selon les analyses, a porté sur des populations de taille variable : complète, 95; fluctuante, 60, 31, 19 selon la participation; échantillonnée, 12, pour certaines analyses relationnelles; sélective, 53, soit les étudiants des groupes I et II avec des dossiers complets.

### 2.2.2 La terminologie opérationnelle

Pour contourner le problème posé par l'absence d'une terminologie conventionnée, les termes suivants ont été utilisés dans l'étude.

« Rubrique » – Ensemble d'éléments d'information, d'ordre similaire, sur un même sujet, réunis dans un même espace informatique, ou page, accessible par un lien hypertexte.

« Clic » – L'équivalent de « hits » avec le sens de « sélection par hypertexte d'un sujet ou d'une rubrique »; ainsi, chacune des entrées dans le registre des *logs* correspond à un clic.

« Connexion » – C'est par un premier clic, spécial mais enregistré comme simple clic, que s'établit une connexion entre un étudiant et l'environnement informatique, un jour donné, à une heure donnée; comme Moodle n'a pas de programme de déconnexion automatique en l'absence de clics, significatifs d'une navigation effective durant une période donnée, ce qui peut donner lieu à des temps de connexion très longs et non significatifs, et comme le « *logout* » n'est pas enregistré comme tel, seule une analyse comparée des temps de navigation permet d'identifier entrées et sorties, donc les connexions. Par ailleurs, le clic signifiant le passage d'une page à une autre, le temps associé à un clic mesure la durée d'arrêt sur une page donnée sélectionnée. Selon la forme d'importation des données de Moodle, le temps est rapporté soit en secondes en temps universel coordonné (UTC) avec le « *Unix Time Stamp* », soit par une chaîne de caractères (*string*), an-mois-jour-heures-minutes-secondes; dans un cas comme dans l'autre, un traitement s'impose pour repérer et mesurer la durée d'une connexion ou d'un clic.

« Itération » – Dans Moodle, chaque nouvelle connexion mène directement à la page principale; dans les *logs*, l'information jumelée « *course view* », sans autre spécification, identifie la page principale; lorsque cette information paraît au cours d'une navigation continue, elle signifie un retour à la page principale; il est donc possible de définir de façon opérationnelle une itération comme la séquence de navigation effectuée entre deux passages par la page principale; Moodle offre, par ailleurs, la possibilité de naviguer sans revenir à la page principale. Les passages par la page principale représentent 30 % des clics enregistrés, ils résultent en majorité du « *back click* », usuel dans la navigation.

Lors des exportations préprogrammées de données, Moodle organise l'information en six (6) grandes catégories de rubriques : *Chat*, *Course*, *Quiz*, *Resource*, *Forum* et *User*; et chacune de celles-ci se subdivise. En relation avec l'étude, il faut signaler que la catégorie *Quiz* englobe le problème à résoudre, les exercices et l'évaluation finale, et la catégorie *Resource*, les rubriques d'instruction, celles des ressources proprement dites et les outils. La catégorie *User* correspond au répertoire avec les fonctions d'inscription, d'actualisation et d'annuaire des adresses de courriel.

### 2.2.3 La procédure suivie

Le traitement analytique des registres de Moodle s'est effectué selon une procédure itérative : identification des variables, importation, nettoyage, enregistrement et/ou transfert, traitement analytique, formatage des résultats, puis reprise, totale ou partielle, du cycle itératif.

#### *Importation des données*

L'importation des données de Moodle s'est effectuée selon diverses modalités au gré de l'analyse : importation du registre des *logs* au complet (17 287 entrées), par rubriques (32 rubriques), par variables, pour 53 étudiants, exportées vers SPSS, et par usagers pour les 12 cas de l'échantillon; ces registres ont tous été nettoyés pour ne conserver que les données concernant les étudiants (8 862 en-

trées). L'échantillonnage s'est fait au hasard à partir de la population sélectionnée de 53 d'entre eux. Selon les analyses, les variables du temps ont été transformées et/ou segmentées pour discriminer une information numérique mesurable et significative. Les données ainsi obtenues ont été enregistrées dans des feuilles de calcul Excel. Trois types d'analyse ont été réalisés.

#### *Analyse de la navigation*

La navigation, *stricto sensu*, de chacun des étudiants n'a pu être traitée à défaut d'une grille de séquences répétitives, catégorisables et catégorisées, résultant d'une analyse séquentielle d'itinéraires suivis dans un espace informatique d'apprentissage. Aussi l'analyse s'est-elle limitée d'abord à décrire pour la population totale quelques-uns des paramètres caractérisant la navigation, soit la fréquence des clics selon les leçons, les rubriques et la participation, et à estimer à l'aide de la population échantillonnée la durée des connexions, des itérations et des clics. L'analyse a ensuite répertorié, pour la population sélectionnée des 53 étudiants, la rubrique immédiatement visitée lors des deux premières connexions subséquentes à l'inscription et à la mise à jour du profil. Cette dernière analyse visait à vérifier si les étudiants abordaient l'espace informatique d'apprentissage d'une même façon et en accord, ou non, avec les instructions fournies. Enfin, un exemple d'analyse séquentielle possible est présenté.

#### *Analyse de l'apprentissage*

Le deuxième type d'analyse a consisté, d'un côté, à vérifier, avec la population totale, si l'usage de l'environnement informatique semblait favoriser l'apprentissage et, de l'autre, à regrouper (*cluster*) les 53 étudiants de la population sélectionnée en fonction de l'amplitude de leur activité dans l'environnement, mesurée par le nombre total de clics, variable indépendante, et les notes obtenues, variables dépendantes, aux examens en ligne et en salle.



L'examen en salle pouvait être considéré, *a priori*, comme indépendant de l'usage ou non de l'environnement informatique, compte tenu des conditions de sa réalisation; celui en ligne, par contre, devait refléter l'usage de cet environnement comme moyen d'apprentissage. L'emploi des deux types d'examen comme variable indépendante devait permettre d'évaluer si différence il y avait. Cette analyse de groupement visait à dégager des classes d'étudiants catégorisées à la fois selon leur activité, définie par le nombre total de clics, et leur apprentissage, défini par les notes obtenues.

#### *Analyse de l'efficacité*

La dernière analyse visait à formuler un indicateur rendant compte de l'efficacité de la navigation définissant l'efficacité sur une base opérationnelle par le nombre de points de qualification académique obtenus aux examens en fonction du nombre total de clics employés lors de la navigation. Cette mesure peut aussi bien se prendre pour un individu que pour un ensemble d'individus et elle varie entre 0 et Q, Q étant la note maximale possible dans un système de notation donné. Pour pouvoir comparer un tel indicateur, il est préférable de le normaliser, soit en fonction d'un maximum possible, soit en fonction de l'efficacité moyenne de la population étudiée. C'est cette dernière option qui a été retenue dans l'étude en appliquant la formule suivante :

$$\epsilon_j = \frac{\sum_{i=1}^n q_{ij}}{\sum_{i=1}^n a_{ij}} \quad / \quad \frac{\sum_{i=1}^n \sum_{j=1}^m q_{ij}}{\sum_{i=1}^n \sum_{j=1}^m a_{ij}}$$

où  $\epsilon$  = l'indice d'efficacité;  
 q = la note à l'examen;  
 a = le nombre de clics;  
 i = les étudiants;  
 j = les groupes d'étudiants;  
 n = le nombre d'étudiants;  
 m = le nombre de groupes;

si  $\epsilon = 1 \Rightarrow$  l'efficacité du groupe est égale à la moyenne générale;

si  $\epsilon < 1 \Rightarrow$  l'efficacité du groupe est inférieure à la moyenne générale;

si  $\epsilon > 1 \Rightarrow$  l'efficacité du groupe est supérieure à la moyenne générale.

### 3. Résultats

Seuls les résultats concernant l'analyse des registres de Moodle sont présentés; ceux qui sont relatifs à la recherche sur les objets d'apprentissage ne le sont pas, sauf pour les caractéristiques des participants, lesquelles auraient pu expliquer les variations observées. Ces résultats sont néanmoins accessibles (Organista, 2007; Organista et Cordero, 2006; Organista et Lavigne, 2006). L'exclusion répond à des exigences pratiques. Cette enquête a montré, entre autres, que les étudiants approuvaient le recours à des objets d'apprentissage pour appuyer leur démarche d'apprentissage, que l'introduction des technologies dans l'enseignement se justifiait, mais qu'au bout du compte, ils préféraient les modalités mixtes.

### 3.1 Les caractéristiques générales des participants

L'analyse des caractéristiques des participants n'a mis en jeu aucune variation significative susceptible d'influer sur les résultats. Le groupe II démontrait un équilibre ( $\pm 50\%$ ) entre les sexes et le groupe I un léger déséquilibre (68 % de femmes). L'état civil des participants était homogène, célibataires à 80 %. Seule différence appréciable mais non significative, 62 % des étudiants du groupe I ne travaillaient pas tandis que la majorité des étudiants du groupe II travaillaient entre 11 et 20 heures par semaine (parfois même plus). Officiellement, tous étaient inscrits à plein temps, la UABC n'acceptant pas d'inscription à temps partiel.

La moyenne cumulative des notes obtenues par les étudiants, tant au collège (cégep, lycée) qu'à l'université, ne présentait pas de différences notables; sur un total de 10, ils avaient obtenu une note moyenne oscillant entre 8,23 comme minimum et 8,79 comme maximum. De même, les caractéristiques socioéconomiques ne démontraient aucune différence significative. Il semble bien que les conditions d'admission et les examens d'entrée homogénéisent de fait la clientèle estudiantine.

Tous les étudiants avaient accès à la technologie informatique et à Internet à l'université; de plus, 86,5 % du groupe I, 70,6 % du groupe II et 94,4 % du groupe III avaient un ordinateur personnel chez eux, la moitié d'entre eux avec un accès à Internet. Pour tous, il s'agissait d'une première expérience de formation en ligne.

### 3.2 L'analyse de la navigation dans l'environnement informatique

Cette partie de l'analyse est provisoire et partielle tant en raison du nombre réduit de cas étudiés, 53 ou 12, que du fait qu'elle ne traite que d'un seul cours et que, pour l'heure, les outils d'analyse à grande échelle manquent pour établir et interpréter des séquences de clics adéquates pour caractériser les itinéraires suivis lors d'une navigation. Ne sont donc abordés que trois aspects : 1) la fréquence et la durée des connexions, des itérations et des clics; 2) la distribution des clics selon les types de rubriques visitées et selon les deux premières connexions; 3) un exemple de codification et de mise en séquences de la navigation d'un étudiant.

#### 3.2.1 La fréquence et la durée des connexions, des itérations et des clics

L'expérience a commencé la dernière semaine de janvier avec une première et unique connexion, très brève, pour s'inscrire. Puis, comme le montre le tableau 1, les étudiants se sont connectés principalement en février et en avril, soit au début et à la fin de la formation en ligne. Durant ces deux mois, en moyenne, les étudiants ont accédé à la plateforme trois fois chaque mois, se sont branchés pour un total de 55 et 40 minutes par mois, et ont navigué à l'aide de 46 et 51 clics. Au total, ils ont pris un peu plus de 1 heure 45 minutes pour surfer au travers des trois leçons avec 109 clics en moyenne.

**Tableau 1.** Distribution et durée des connexions par étudiant et par mois

N = 12	Mois					Total
	Janvier	Février	Mars	Avril	Mai	
Nombre moyen de connexions	1	3,3	1	2,7	0,2	8,2
Durée (h:m:s) des connexions	00:00:04	00:55:08	00:10:39	00:40:18	00:00:41	01:46:58
Nombre moyen de clics	3	46	6	51	3	109

Cette information d'ordre général, si elle se confirmait avec d'autres analyses, établit un rythme, disons bimodal, au processus d'apprentissage qui mériterait d'être approfondi pour être bien interprété. Résulte-t-il d'une fatigue, comparable à celle des étudiants à distance, à mi-chemin de leur formation? D'une surcharge conjoncturelle de tâches? D'un laisser-aller? Selon la nature de la conclusion, les conséquences seraient différentes sur la facture de la formation.

Le clic, l'itération et la connexion sont les trois unités retenues pour structurer, *a posteriori*, la navigation des étudiants. Aussi l'ampleur et la durée de ces variables donnent-elles une image de la structure moyenne générale des pratiques des étudiants, comme le montre le tableau 2.

**Tableau 2.** Ampleur et durée des unités de navigation des étudiants

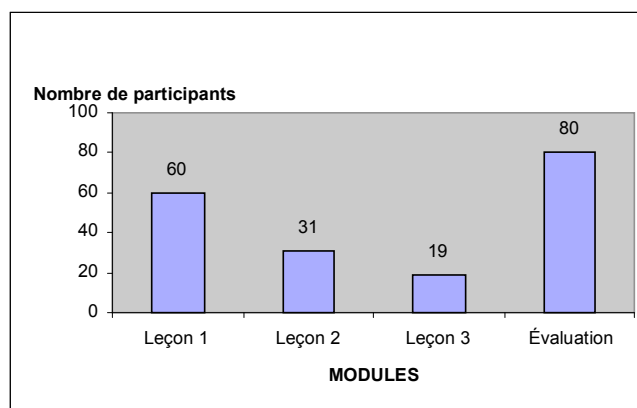
Nombre moyen par étudiant (n = 12)	
Connexions	8,17
Itérations par connexion	3,01
Clics par itération	4,41
Durée moyenne par unité en minutes	
de chaque clic	0:59
de chaque itération	4:19
de chaque connexion	13:05

Par manque de comparaison avec des études similaires, il est difficile d'interpréter ces mesures. Sont-elles particulières aux étudiants mexicains? À la matière? À la formation en ligne? À la modalité mixte? Il est clair, par contre, que dans ce cas précis, la relation entre l'étudiant et l'espace informatique a été dynamique, rapide d'exécution et brève.

Des trois unités, l'itinéraire, la seule forgée, est celle qui démontre une corrélation significative avec les deux autres (connexion :  $r = 0,673$ ;  $0,160 < ,05$ ; clic :  $r = 0,744$ ;  $0,006 < ,01$ ). Cette signification ne fait que souligner l'importance du « *back click* » dans la structuration des itinéraires de navigation. Étant donné, par contre, la fonction de sas jouée par ce clic spécial, rien, de prime abord, ne milite pour sa prise en charge dans une analyse exhaustive des itinéraires de navigation, au contraire, sauf son rôle d'unité intermédiaire de structuration.

### 3.2.2 La distribution des clics

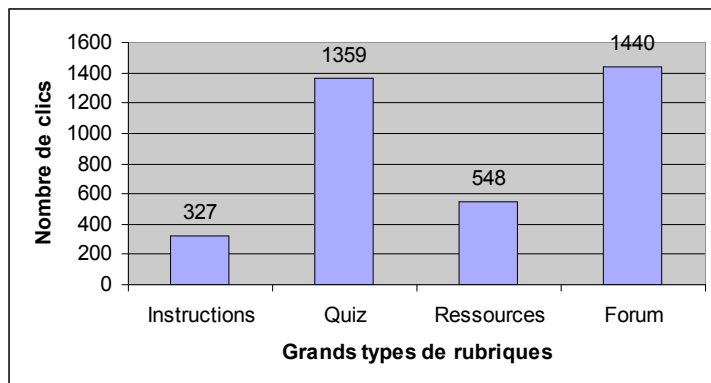
Une donnée, à première vue occultée en raison des conditions mêmes de l'extraction préprogrammée de données, concerne la participation effective des étudiants aux leçons (il faut configurer soi-même le rapport). La figure 1 montre que celle-ci a nettement diminué de leçon en leçon (à noter que ces analyses ont porté sur la population totale). Certes, cette diminution s'explique par la non-participation des étudiants du groupe III, mais celle-ci ne peut rendre compte de la tendance généralisée.



**Figure 1.** Participation des étudiants par leçon et pour l'évaluation

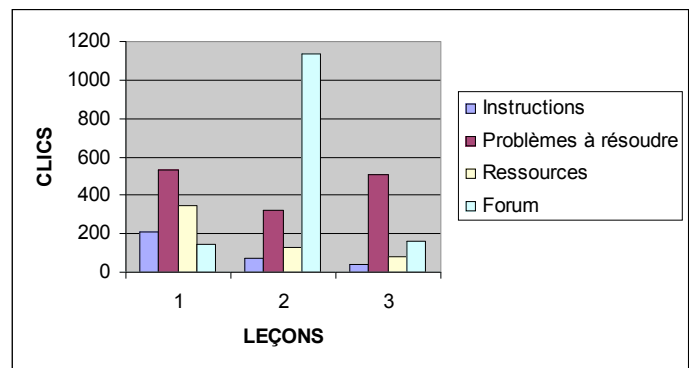
Là encore, l'interprétation est évasive. Doit-on associer la déperdition avec le niveau de difficulté de la matière? Les résultats aux examens partiels en salle contredisent une telle approche; en effet, la note moyenne obtenue varie de 80,3 % pour la courbe normale (avec  $n = 94$ ) à 87,6 % pour la formulation d'hypothèse (avec  $n = 64$ ) et 85,2 % pour les distributions (avec  $n = 64$ ). Tout au plus pourrait-on concevoir l'inverse et associer la déperdition avec la facilité de la matière, ce qui reste discutable. Une conclusion inévitable s'impose néanmoins. Comme tous les étudiants ont participé aux examens en salle, obligatoires car officiels, et qu'ils ont obtenu une note moyenne relativement bonne, la faible participation à la leçon 3 en ligne pourrait signifier que les étudiants se sont reportés sur les leçons en classe, en présentiel donc, pour fonder leur apprentissage. Rien dans les groupes de discussion ne laissait présager une telle attitude. Il serait donc utile d'en rechercher les causes puisque la qualité et/ou la facture du design pédagogique de cette leçon sont possiblement en cause.

La figure 2 illustre la distribution des clics entre les grandes catégories de rubriques, soit celles qui sont relatives aux instructions, aux exercices et à l'évaluation (quiz), aux ressources pédagogiques et aux forums. N'apparaissent ni les rubriques associées au *chat* ni celles qui sont liées au courrier électronique.



**Figure 2.** Distribution des clics par grandes rubriques

De prime abord, cette distribution pourrait paraître surprenante. La compréhension de la logique didactique du design pédagogique voudrait que les rubriques de ressources soient pour le moins autant consultées que les exercices et les forums. Cette distribution, toutefois, peut s'expliquer. Pour les forums, le recours à cette forme de communication était fortement recommandé pour la leçon 2, d'où l'importance prise par cette rubrique, comme le montre la figure 3.



**Figure 3.** Total des clics par leçon et par rubrique

Par ailleurs, l'ampleur des exercices peut également s'expliquer. Les étudiants devaient vérifier leur compréhension de la matière en réalisant les exercices prévus à cette fin, exercices qu'ils pouvaient refaire plusieurs fois, et jusqu'à l'évaluation qu'ils pouvaient reprendre trois fois. On conçoit donc qu'ils aient visité ces rubriques souvent afin de confirmer leur apprentissage. Une interprétation s'impose. L'inclusion de tâches interactives comme moyen d'asserter l'apprentissage et celle de forums comme modalité de dynamisation d'un apprentissage collaboratif semblent bien constituer des composantes inévitables du design pédagogique à en juger par l'usage que les étudiants en ont fait dans cette expérience.

Toutefois, seule une analyse fine des itinéraires de navigation pourrait rendre complètement intelligible la distribution observée. Pour l'heure, par contre, celle-ci échappe par manque d'une définition et d'une instrumentation appropriées. Afin, néanmoins, d'approximer l'attitude initiale des étudiants face à la plateforme, l'analyse a recherché quelle était la rubrique visée lors des deux premières connexions ayant une portée pédagogique (soit les connexions trois et quatre, les deux premières étant réservées à l'inscription et à la mise à jour du profil). Le traitement a porté sur la population sélectionnée de 53 cas et n'ont été utilisées que trois grandes catégories, les ressources, incluant les instructions en plus des ressources didactiques, les quiz, exercices et évaluation, et la communication, incluant forums et courriels.

Comme le montre le tableau 3, la majorité des étudiants ont visité les rubriques de ressources lors de leur première connexion, puis celles de quiz lors de leur deuxième visite.

**Tableau 3.** Distribution des clics selon les grandes rubriques lors des deux premières connexions à portée pédagogique

Connexion (N = 53)	Type de rubrique			Total
	R	Q	C	
Première	34	10	9	53
%	64,15	18,86	16,98	
Deuxième	17	24	10	51
%	33,33	47,05	19,60	

R = Ressources;      Q = Exercices et évaluation;      C = Communication (forums et courriels)

Afin de discerner d'éventuelles stratégies de navigation, différenciées par des séquences types d'accès aux rubriques, neuf (9) classes de catégories jumelées ont été créées.

**Tableau 4.** Distribution des stratégies (accès jumelés) lors des deux premières connexions à portée pédagogique

CLASSES	N	%	% CUMULATIF
RQ	16	30,2	
RR	13	24,5	
RC	4	7,5	62,2
QR	4	7,5	
QC	3	5,7	
QQ	2	3,8	17,0
CQ	6	11,3	
CC	3	5,7	17,0
QØ/RØ	2	3,8	3,8
TOTAL	53		

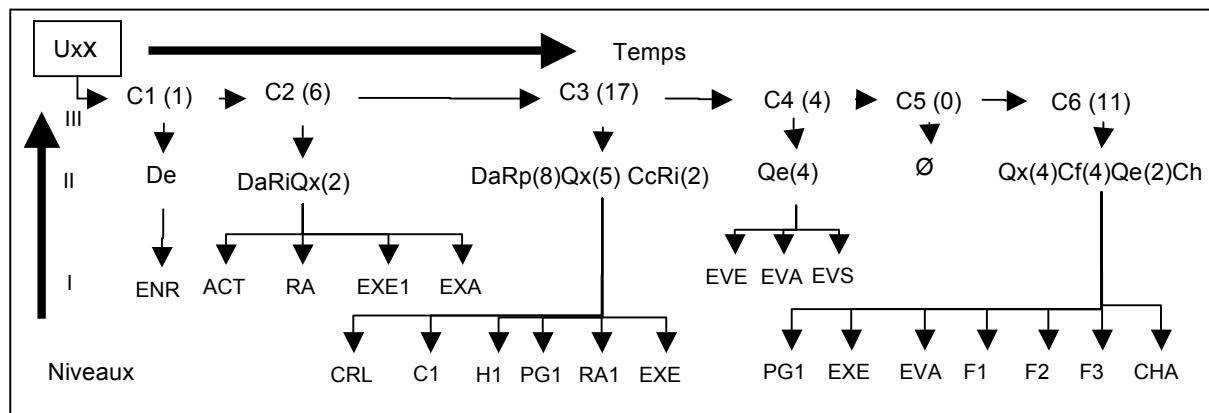
R = Ressources    Q = Exercices et évaluation    C = Communication

D'après le tableau 4, ce sont bien trois types différents de navigation initiale qui ressortent. Le premier, majoritaire, caractérise les étudiants qui entrent dans l'espace informatique par une rubrique de ressources, le deuxième par les quiz et le troisième par la communication. Les deux premiers types réorientent la navigation lors de la deuxième connexion vers les autres grandes rubriques, alors que le troisième type ignore les rubriques de ressources. À noter, deux cas d'absence d'orientation.

Cette analyse laisse présager l'existence de formes d'accès différentes à l'espace informatique : une première à l'aide des ressources, la plus importante et la plus prévisible compte tenu du design pédagogique, une deuxième par les exercices et une troisième par les communications. Alors que les deux premières semblent renvoyer à des styles de navigation favorisant l'apprentissage en fonction d'une prise d'information puis d'une mise en application, la dernière semble s'inscrire dans une perspective autre où « la connaissance, c'est le travail d'équipe », tel que l'ont discutée Dziuban *et al.* (2005). Enfin, les deux cas d'absence de mouvement informent de l'existence d'un problème, pédagogique, didactique ou technique, information facile à signaler par une programmation appropriée.

Pour à la fois saisir la complexité et la richesse d'une telle analyse, voici un exemple de traitement de la navigation d'un étudiant. Celui-ci s'est connecté six fois entre janvier et avril 2006 et il a navigué à l'aide de 70 clics. Pour catégoriser les pas de sa navigation, 21 codes ont été créés, plus 12 autres pour mettre en séquences ces mêmes pas<sup>2</sup>. Comme le montre la figure 4, la codification et la mise en séquences des clics se traduisent en une structure sur trois niveaux.

Au niveau I, la codification qualifie les clics qui informent de l'itinéraire suivi, pas à pas; au niveau II, les séquences itératives informent de l'ampleur et de l'assemblage relationnels des itérations; au niveau III, les connexions, si elles étaient portées



**Figure 4.** Exemple de structuration relationnelle d'une navigation

### 3.2.3 L'exemple d'analyse séquentielle

L'analyse détaillée des itinéraires de navigation des étudiants se compose dans un premier temps d'une étape de codification (définition et application) de chacun des clics qui balisent l'itinéraire suivi par un étudiant, puis d'une seconde de mise en séquences de ces codes pour construire des unités récursives et possiblement significatives dont les agencements morphologiques seraient caractéristiques de composantes de navigation.

sur une échelle temporelle, schématiseraient la relation entre l'étudiant et l'environnement d'apprentissage. La démarche vise donc à condenser l'information sans perdre sa qualité relationnelle.

La lecture, par contre, se fait du niveau III vers le premier. Ainsi, cet étudiant, après s'être inscrit (C1) (où C1 = connexion # 1 et le chiffre entre parenthèses indique le nombre de clics sans compter les pages principales), a actualisé (C2) son profil (Da) (le nombre de clics n'apparaît entre parenthèses qu'à partir de 2), s'est informé du travail à faire (Ri) et a vérifié sa compétence à réaliser les exercices (Qx). Il a ensuite entrepris (C3) une série

d'itérations entre ressources et exercices (Rp-Qx), après avoir actualisé son profil (Da), communiqué avec un collègue (Cc) et revu les instructions (Ri). À mi-chemin (C4), il a jaugé sa compétence avec l'évaluation (Qe). Une connexion vide, inexplicable. Pour terminer sa navigation, il a repris (C6) les exercices (Qx), vérifié les forums (Cf) (sans y participer), *chatté* avec des collègues (Ch) pour ensuite vérifier son évaluation finale (Qe). Sa navigation n'a porté que sur la leçon 1 (signifié par le chiffre 1 au niveau I), peut-être s'agissait-il d'un étudiant du groupe III. Le temps n'a été pris en compte que pour signifier l'ordre d'occurrence des séquences et non dans sa dimension durée.

Cette analyse, relationnelle, met en œuvre une méthodologie présentant des similitudes avec l'analyse linguistique, voire génétique. La comparaison entre les structures de navigation de plusieurs étudiants permettrait de distinguer des segments récurrents de séquences, possiblement significatives de types de navigation, selon leurs règles d'association. La mise au jour de tels profils de navigation faciliterait, par l'intermédiaire du design, l'adaptation des composantes d'un environnement virtuel d'apprentissage aux caractéristiques, sinon des individus, du moins d'ensembles d'étudiants ayant un même profil, option plus réaliste dans le cadre de la formation de masse.

### 3.3 L'analyse de la relation environnement-apprentissage

L'apprentissage, comme tel, est inaccessible, seules les traces qu'il laisse sont éventuellement perceptibles. La notation d'exercices et d'examens n'est certes pas la meilleure façon d'approximer l'apprentissage, mais c'est la plus courante. Afin d'évaluer l'impact que l'immersion dans l'environnement d'apprentissage pouvait avoir eu sur l'apprentissage des étudiants, trois analyses ont été menées. Une première visait à discerner un éventuel gain de performance aux exercices, à l'évaluation en ligne et à l'examen en salle. Une deuxième visait à vérifier si le regroupement des étudiants en fonction de leur niveau d'activité

dans l'environnement et des notes obtenues aux tests en ligne et en salle était significatif. Enfin, une troisième analyse visait à explorer la possibilité de créer un indicateur de l'efficacité pédagogique des navigations selon les groupes obtenus.

#### 3.3.1 L'analyse de la performance aux exercices et aux examens

Comme le montre le tableau 5, si au début de la formation les étudiants ont, dans l'ensemble, obtenu un pointage plutôt faible aux exercices, ils ont, au moment de l'évaluation en ligne, obtenu des notes nettement meilleures; quant à la note finale, c'est-à-dire la moyenne des trois examens partiels réalisés en classe sous la direction du professeur responsable, elle est congruente avec la tendance observable d'une amélioration.

**Tableau 5.** Distribution des notes moyennes sur 100 obtenues aux exercices, à l'évaluation en ligne et à l'examen en classe

Quiz N = 95	Notes sur 100
Exercices	22,73
Évaluation (en ligne)	65,89
Examen (en classe)	83,20

Il est toutefois difficile d'associer d'emblée le gain perceptible avec l'immersion dans l'environnement informatique d'apprentissage. Les étudiants savaient que les résultats aux exercices ne seraient pas comptabilisés dans leur évaluation officielle. De plus, l'évaluation en ligne n'a pas non plus été prise en compte, seules les notes obtenues aux examens en salle ont été officialisées.

### 3.3.2 L'analyse de la relation entre l'activité et l'évaluation

Le processus d'apprentissage est individuel, biologiquement fondé. Mais l'apprentissage dépend d'un ensemble de facteurs dont des connaissances préalables et le contexte dans lequel il prend forme. La composition du contexte, selon l'angle de sa perception, peut être très vaste, diversifiée et spécifique par certains aspects. Les composantes de ce contexte les plus étudiées sont les « autres » et l'« environnement immédiat d'apprentissage ». Les étudiants engagés dans l'expérience avaient tous une formation préalable équivalente en statistique. Pour tous, l'environnement informatique constituait un ajout particulier et nouveau à leur contexte d'apprentissage.

La situation éducative dans laquelle les étudiants se trouvaient autorisait donc de s'interroger sur l'existence d'une éventuelle relation entre le degré d'activité dans l'environnement et les notes obtenues. Le traitement analytique le plus simple et le plus direct pour cerner ce type de relation est l'analyse de regroupement (*cluster analysis*). Pour ce faire, l'activité a été mesurée par le nombre total de clics, variable indépendante, et les notes obtenues, variable dépendante, à l'évaluation en ligne et aux examens en salle (moyenne des trois). La population traitée a été celle des 53 étudiants sélectionnés. Le tableau 6 montre que le test F, dans les deux cas, dénote l'existence de quatre groupes distincts de façon significative. Au bout du compte, ce sont cinq groupes que cette analyse a discriminés. Le traitement, en effet, a exclu de l'analyse les quatre cas extrêmes d'activité, lesquels ont été regroupés dans un cinquième groupe.

**Tableau 6.** Regroupement des cas en quatre groupes selon l'activité, l'évaluation en ligne et l'examen en classe

ANOVA	Groupes		Erreur		F	Sig.
	Moyenne <sup>2</sup>	dl	Moyenne <sup>2</sup>	dl		
1 <sup>er</sup> regroupement						
Total des clics	28265,149	3	225,494	45	125,347	,000
Examen (en classe)	1642,594	3	101,965	45	16,109	,000
2 <sup>e</sup> regroupement						
Total des clics	27989,308	3	243,884	45	114,765	,000
Évaluation (en ligne)	2054,817	3	250,546	45	8,201	,000

Le dénombrement, cas par cas, des étudiants selon les deux regroupements résultant de l'analyse a montré que les groupes se chevauchaient largement. Le groupe « 1 en classe » recouvrait 82,35 % des étudiants des groupes 1 et 2 « en ligne »; les groupes « 3 en classe » et « 3 en ligne » coïncidaient parfaitement; 82,61 % des étudiants des groupes « 2 en classe » et « 4 en ligne » étaient les mêmes. Par contre, le groupe « 4 en classe » n'avait pas d'équivalent dans le regroupement en ligne puisqu'il se compose de cas issus de 3 groupes « en ligne » différents; ce groupe mobilise 7,5 % des effectifs. Ceci implique que l'un ou l'autre type de notation peut être utilisé sans que les groupes qui en résultent diffèrent suffisamment pour justifier de doubler l'analyse.



### 3.3.3 L'analyse de l'efficacité pédagogique relative de la navigation

La question posée est simple : existe-t-il une relation entre l'activité dans l'espace informatique et l'apprentissage telle que plus grande l'activité, meilleur l'apprentissage? La réponse est ambiguë. Il n'y a pas de corrélation significative sur le plan statistique entre les deux variables. Par contre, les étudiants se regroupent bel et bien en classes définies en fonction de leur niveau d'activité et de leurs notes, ceci de façon significative. Autrement dit, oui, il existe une relation, mais elle doit être modulée. Ainsi les étudiants avec une note excellente se partagent-ils entre ceux qui ont un haut niveau d'activité et ceux qui ont un bas niveau, chacun constituant un groupe distinct. Idem pour les étudiants ayant une note moyenne ou médiocre. Ce constat ouvre la porte à une définition opérationnelle de l'efficacité pédagogique de la navigation qui permet de définir des catégories sur la base de ces deux variables, à savoir : combien faut-il de clics pour obtenir un point de notation, ou à l'inverse combien de points de notation obtient-on par clic?

La mesure utilisée a été celle présentée plus haut dans le protocole. Le tableau 7 présente les quatre types d'efficacité mesurée de façon relative à l'efficacité moyenne de l'ensemble des étudiants.

**Tableau 7.** Efficacité relative de la navigation par groupes

Efficacité relative de la navigation (ε)				
Type d'efficacité	Groupes issus de l'analyse de <i>cluster</i>	Type de notation		% d'étudiants
		En classe	En ligne	
Très faible	5	0,34		7,5
	5		0,36	
Faible	3	0,64		13,2
	3		0,63	
Moyenne	1	1,06		35,8
	4	0,95		
	<i>Moyenne</i>	<i>1,01</i>		
	1		1,21	
	2		0,67	
	<i>Moyenne</i>		<i>0,94</i>	
Forte	2	2,35		43,4
	4		2,26	

Première constatation : la valeur de l'indice varie très peu d'un type de notation à l'autre. Seconde constatation : l'indice calculé pour chacun des groupes discriminés par l'analyse de regroupement démontre une variation suffisante pour les démarquer.

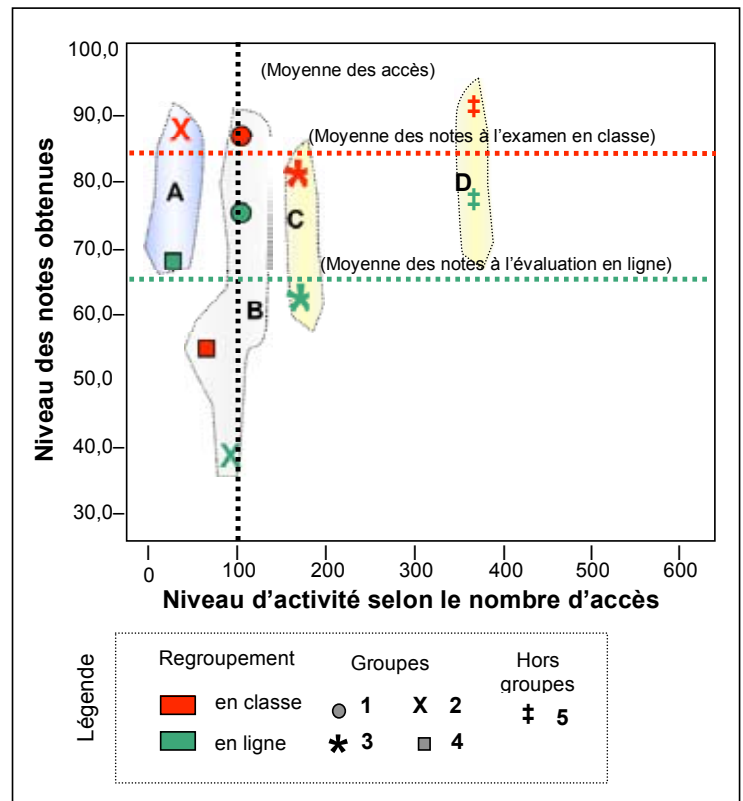
Le groupe 5 (ensemble D dans la figure 5 à la page suivante) réunit les étudiants, exclus du calcul de regroupement, ayant démontré un niveau d'activité extrême (entre 400 et 550 clics); ils sont caractérisés par une très faible efficacité même si les notes obtenues étaient bonnes, voire excellentes pour certains. L'indice exprime donc une particularité de navigation, possiblement assimilable à une difficulté, pédagogique ou technique, qui a requis de la part des étudiants un investissement en temps nettement supérieur à la normale, telle que signifiée par les données d'une navigation moyenne. Un tel groupe mériterait peut-être une attention et un support appropriés, mais il faudrait d'abord approfondir les motifs d'une telle suractivité.

Les étudiants du groupe 3 (ensemble C dans la figure 5) démontrent une efficacité faible découlant à la fois d'une forte activité et d'une performance d'apprentissage médiocre. L'indice signale par conséquent un groupe d'étudiants en difficulté d'apprentissage; ils travaillent beaucoup, mais sans en tirer les profits escomptables. Il est plausible, dans ce cas, que la difficulté soit liée à la matière.

Le regroupement caractérisé par une efficacité moyenne réunit le tiers des étudiants dont les notes s'échelonnent de part et d'autre de la moyenne, de bonnes à mauvaises, pour un niveau d'activité moyen. Sur le plan de l'apprentissage, ce groupe est composite; il regroupe autant d'étudiants sans réels problèmes d'apprentissage que d'étudiants en difficulté. L'indice est ici biaisé. Il serait donc souhaitable de regrouper les étudiants selon la classification générée avec la notation de l'évaluation en ligne. En effet, les groupes qui en résultent présentent des indices plus tranchés.

L'analyse du dernier groupe (ensemble A) laisse entendre que lors de l'expérience, la majorité des étudiants a obtenu de bonnes notes avec une navigation d'ampleur limitée nettement plus efficace que celle de leurs collègues. Ce groupe, évidemment, ne pose aucun problème.

Cette analyse de l'efficacité de la navigation n'a de valeur qu'exploratoire. Elle laisse néanmoins entrevoir l'à-propos de développer un tel indice compte tenu des possibilités qu'il pourrait offrir de distinguer en cours d'apprentissage, de façon automatique sur une base probabiliste, les étudiants en difficulté tant pédagogique que technique. Pour ce faire, il serait nécessaire d'analyser plus en détail les itinéraires suivis sur une base temporelle dans le but de dégager des caractéristiques associées aux divers types d'efficacité afin de modéliser ces cheminements en début d'apprentissage et d'établir des seuils de signification.



**Figure 5.** Distribution des groupes en fonction des notes obtenues et du niveau d'activité dans l'environnement d'apprentissage

## Conclusion

L'analyse réalisée est préliminaire et les résultats obtenus provisoires. Ceux-ci mettent en jeu un type de connaissances particulier, la statistique, et ce, dans un pays en particulier, le Mexique; ils ne peuvent donc se généraliser d'emblée. Par ailleurs, la modalité d'apprentissage, présentiel/virtuel, pose la question de l'influence combinée de ces modalités sur le processus, c'est là un objet de recherche en soi (Backhoff *et al.*, 2007; Lavigne *et al.*, 2006, 2008). Par contre, les analyses du registre des *logs* délimitent un espace de recherche riche d'une information potentiellement opportune pour comprendre, suivre et appuyer le cheminement d'apprentissage des étudiants dans un environnement informatique.

Les résultats ne permettent pas de valider l'existence d'une relation directe, significative au sens de la statistique, entre l'utilisation des ressources de l'environnement informatique et l'apprentissage réalisé. Ils montrent, par contre, une tendance dans ce sens. Ainsi, le regroupement des étudiants selon ces deux variables délimite des catégories d'étudiants caractérisées par une efficacité relative de leur navigation en ce qui concerne l'apprentissage, efficacité mesurable et interprétable selon les difficultés pour plusieurs groupes.

Même si la taille restreinte des populations utilisées en limite la portée, l'analyse des pratiques de navigation des étudiants met en jeu des caractéristiques intéressantes. Elle montre le libre arbitre dont ces derniers font preuve dans leur utilisation de l'environnement d'apprentissage quant au cheminement à suivre et à l'implication temporelle consentie par chacun d'eux. Elle montre également l'importance des activités interactives dotées d'une rétroalimentation corrective, tels les exercices dans ces cours, et des moyens de communication entre pairs, courriels, forums, *chat*, comme leviers d'apprentissage.

L'étude met en œuvre des variables caractérisant des stratégies d'apprentissage. La durée, la fréquence et la séquence des connexions, des itérations et des clics, ou encore les accès initiaux, toutes ces variables contribuent à la construction d'une morphologie d'itinéraires de navigation que des études plus poussées permettront de différencier selon des catégories d'étudiants internautes. Les pratiques de navigation des étudiants s'inscrivent dans la mouvance de la nouvelle dynamique techno-pédagogique à la source de nouveaux modes d'apprentissage, encore non formalisés.

L'analyse des *logs* d'environnements d'apprentissage informatiques délimite un terrain de recherche encore peu exploité. Les résultats obtenus, en effet, révèlent des coins d'ombre dont la connaissance demande qu'ils soient explorés :

- Reprise des analyses pour vérifier l'existence ou non de différences dans les pratiques de navigation des étudiants selon les niveaux d'efficacité établis;
- Conduite d'analyses séquentielles et comparées des navigations pour en dégager des styles;
- Confrontation des styles d'apprentissage et de navigation;
- Analyse de l'effet du temps sur les pratiques de navigation.

Par ailleurs, la généralisation de la modalité hybride demande que les caractéristiques de la formation virtuelle soient bien établies, comprises et contrôlées pour que cette modalité donne les bénéfices qui lui sont attribués *a priori* (Lavigne *et al.*, 2008), ce qui implique, entre autres, une révision critique du processus de design (Lavigne, 2007). L'échec relatif du « *e-learning boom* » au début des années 2000 aux États-Unis (Zemsky et Massy, 2004) a souligné l'à-propos de l'approfondissement des tenants et aboutissants de cette modalité d'apprentissage.

La virtualisation du processus d'enseignement-apprentissage génère des sommes énormes de données, fortes d'une information parfois insoupçonnable, mais desquelles découle une surcharge d'information, inévitable. Les *logs* d'une plateforme comme Moodle recèlent bel et bien une information à portée pédagogique importante, qu'il serait téméraire d'ignorer, mais dont le traitement analytique reste laborieux. La recherche du sens de l'information contenue dans ces bases de données chemine entre la fouille de données, méthode de traitement on ne peut plus heuristique, et l'analyse exploratoire des données, méthode de traitement systématique. Dans un cas comme dans l'autre, la quête de la connaissance navigue de l'utopie vers une réalité, partie mouvante, partie évanescence, toujours à découvrir.

## Références

- Backhoff, E. E., Lavigne, G., Organista, J. S. et Aguirre Muñoz, L. C. (2007). Modelo mixto de educación digital-presencial: Una alternativa para el posgrado en México. Dans M. G. Cárdenas López, A. Vite Sierra et L. Villanueva (dir.), *Ambientes virtuales para la educación y para la rehabilitación psicológica* (p. 17-33). Mexico : Universidad Nacional Autónoma de México.
- Bisson, G., Bronner, A., Gordon, M. B., Nicaud, J.-F. et Renaudie, D. (2003, avril). Analyse statistique de comportements d'élèves en algèbre. Dans C. Desmoulins, P. Marquet et D. Bouhineau (dir.), *EIAH2003 Environnements informatiques pour l'apprentissage humain. Actes de la conférence EIAH 2003* (p. 67-78). Paris : Institut National de Recherche Pédagogique [INRP]. Récupéré du site des archives du colloque : <http://archive.eiah.univ-lemans.fr/EIAH2003>
- Bisson, G. et Merceron, A. (2007, juillet). *Fouiller les traces d'activité d'apprentissage*. Communication présentée à la 5<sup>e</sup> école thématique du CNRS sur les EIAH – Personnalisation des EIAH, Saint-Quentin-sur-Isère, France. Résumé récupéré du site de l'école thématique, section *Programme* : [http://eiah2007.imag.fr/bisson\\_merceron.htm](http://eiah2007.imag.fr/bisson_merceron.htm)
- Buelens, H., Roosels, W., Wils, A. et Van Rentergem, L. (2002). 26 One year E-learning at the K.U. Leuven: An examination of log-files. Dans A. J. Kallenberg et M. J. J. M. van de Ven (dir.), *The New Educational Benefits of ICT in Higher Education: Proceedings* (p. 170-174). Rotterdam, Pays-Bas : Erasmus Plus bv/OEER. Récupéré de : <http://publishing.eur.nl/ir/repub/asset/1248>
- Charlier, B., Deschryver, N. et Peraya, D. (sous presse). Face-to-face and distance learning: How to define hybrid systems? How to account for them? *International Journal of Networking and Virtual Organizations*.
- Cheyne, A. (2006, mai). *Recherche de motifs séquentiels pour guider l'interprétation des traces d'apprentissage dans un EIAH*. Communication présentée aux 1<sup>res</sup> Rencontres Jeunes Chercheurs en EIAH, RJC-EIAH, Every, France.

- Daedalus – Data, Decisions and Language, S. A. (2008). *Minería de datos* [Data mining]. Récupéré le 6 août 2008 du site de l'entreprise : <http://www.daedalus.es/mineria-de-datos>
- de Baker, R. S. J., Barnes, T. et Beck, J. E. (dir.). (2008). *Educational Data Mining 2008: 1<sup>st</sup> International Conference on Educational Data Mining, Proceedings*. Montréal, Canada : International Working Group on Educational Data Mining.
- Durán, E. et Costaguta, R. (2007, 10 mars). Minería de datos para descubrir estilos de aprendizaje. *Revista Iberoamericana de Educación – De Los Letores*, 42(2). Récupéré du site de la revue, section *De los lectores – Investigación Educativa* : <http://www.rieoei.org/1674.htm>
- Dziuban, C., Moskal, P. et Hartman, J. (2004). Higher education, blended learning and the generations: Knowledge is power – no more [version électronique]. Dans J. Bourne et J. C. Moore (dir.), *Elements of quality online education: Engaging communities*. Needham, MA : Sloan Center for Online Education. Récupéré du site du Teaching & Learning Center de l'Université de Calgary, section *Stories – eLearning The Millenial Generation* : <http://tlc.ucalgary.ca/documents/chuck.doc>
- Eyssautier-Bavay, C. et Ollagnier-Beldame, M. (2006, mai). *Médiatiser la prise de conscience métacognitive en contextes d'apprentissage*. Communication présentée aux 1<sup>res</sup> Rencontres Jeunes Chercheurs en EIAH, RJC-EIAH, Every, France. Récupéré du site de l'auteur à l'Université Lyon 1 : <http://www710.univ-lyon1.fr/~ceyssaut/publications>
- Feng, M. et Heffernan, N. T. (2005). Informing teachers live about student learning: Reporting in the assistent system [version électronique, pré-publication]. *Technology, Instruction, Cognition, and Learning*, 3(1-2), 63-72. Récupéré le 5 février 2007 du site de l'auteur au Worcester Polytechnic Institute : <http://web.cs.wpi.edu/~mfeng>
- Feng, M. et Heffernan, N. T. (2007, avril). *Assessing students' performance longitudinally : Item difficulty parameter vs. skill learning tracking*. Communication présentée au National Council on Educational Measurement 2007 Annual Meeting, Chicago, IL. Récupéré du site de l'auteur au Worcester Polytechnic Institute : <http://web.cs.wpi.edu/~mfeng>
- Ferraris, C. et Lejeune, A. (2007, juillet). *Le couplage « scénario – traces »*. Exemple du langage LDL. Communication présentée à la 5<sup>e</sup> école thématique du CNRS sur les EIAH – Personnalisation des EIAH. Saint-Quentin-sur-Isère, France. Résumé récupéré du site de l'école thématique, section *Programme* : [http://eiah2007.imag.fr/ferrari\\_lejeune.htm](http://eiah2007.imag.fr/ferrari_lejeune.htm)
- Fischer, F. (2005). *L'utilisation des traces numériques dans l'enseignement à distance*. Rapport de recherche [pour fins d'obtention d'une maîtrise] non publié, École nationale supérieure des sciences de l'information et des bibliothèques, Villeurbanne, France. Récupéré du site du directeur de recherche à l'Institut d'informatique et mathématique appliquées de Grenoble, section *Recherche – Encadrement* : <http://www-clips.imag.fr/arcade/User/jean-philippe.pernin/recherche/encadrement.htm>
- Hardy, J., Bates, S., Antonioletti, M. et Seed, T. (2005). Integrating e-learning and on-campus teaching II: Evaluation of student use [version électronique]. Dans J. Cook et D. Whitelock (dir.), *Exploring the frontiers of e-learning: Borders, outposts and migration. Research proceedings of the 12<sup>th</sup> Association of Learning Technology Conference* (p. 140-153). Oxford : Association for Learning Technology. Récupéré le 12 février 2007 du site de la School of Physics and Astronomy de l'Université d'Édimbourg, section *e-Learning in Physics – Publications* : <http://www.ph.ed.ac.uk/elearning/publications>
- Lancieri, L. (2007, juillet). *De l'analyse de traces à l'exploitation des phénomènes d'intelligence collective*. Communication présentée à la 5<sup>e</sup> école thématique du CNRS sur les EIAH – Personnalisation des EIAH. Saint-Quentin-sur-Isère, France. Résumé récupéré du site de l'école thématique, section *Programme* : <http://eiah2007.imag.fr/lancieri.htm>

- Lavigne, G. (2007). La pédagogie contemporaine de l'apprentissage constructiviste : autoroute ou cul de sac pour la formation à distance? *Revue de l'éducation à distance*, 22(1), 59-72. Récupéré du site de la revue, section Archives : <http://www.jofde.ca/index.php/jde/issue/view/2>
- Lavigne, G., Backhoff-Escudero, E. et Organista, J. S. (2008). La hibridación digital del proceso educativo. Dans J. Vales Garcia (dir.), *Nuevas tecnologías para el aprendizaje* (p. 43-59). Mexico : Person / Prentice Hall.
- Lavigne, G., Organista, J. S. et Aguirre Muñoz, L. C. (2006). Evaluación de la modalidad híbrida, presencial/en línea, por estudiantes de posgrado en educación. *Revista Electrónica Actualidades Investigativas en Educación*, 6(1). Récupéré du site de la revue, section Ediciones : <http://revista.inie.ucr.ac.cr/articulos/1-2006/evaluacion.php>
- Merceron, A. et Yacef, K. (2005). Educational data mining: A case study [version électronique]. Dans C.-K. Looi, G. Mccalla, B. Bredeweg et J. Breuker (dir.), *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology* [Papers presented at the 12th International Conference on Artificial Intelligence in Education] (p. 467-474). Amsterdam, Pays-Bas : IOS Press. Récupéré le 16 septembre 2008 du site de l'auteur à l'Université de Sydney, section Publications : <http://www.it.usyd.edu.au/%7Ekalina/publish.html>
- Merceron, A. et Yacef, K. (2008). Interestingness measures for association rules in educational data. Dans R. S. J. de Baker, T. Barnes et J. E. Beck (dir.), *Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings* (p. 57-66). Montréal, Canada : International Working Group on Educational Data Mining. Récupéré du site de la conférence, section Proceedings : [http://www.educationaldatamining.org/EDM2008/uploads/proc/6\\_Yacef\\_18.pdf](http://www.educationaldatamining.org/EDM2008/uploads/proc/6_Yacef_18.pdf)
- Organista, J. S. (2007). *Desarrollo y evaluación de lecciones en línea para la enseñanza de Estadística basadas en el constructivismo y objetos de aprendizaje*. Thèse de doctorat non publiée, Universidad Autónoma de Baja California, Ensenada, Mexique.
- Organista, J. S. et Cordero, G. A. (2006). Estadística y objetos de aprendizaje. Una experiencia in vivo. *Revista Apertura*. 6(5), 22-35.
- Organista, J. et Lavigne, G. (2006). Desarrollo y aplicación de lecciones de estadística en-línea con objetos de aprendizaje en un ambiente universitario. *Revista Electrónica en Actualidades Investigativas en Educación*, 6(3). Récupéré du site de la revue, section Ediciones : <http://revista.inie.ucr.ac.cr/articulos/3-2006/estadistica.php>
- Organista, J. S., Lavigne, G. et McAnally, L. S. (sous presse). Aprendizaje de estadística y la actividad en línea del estudiante: estudio preliminar de relaciones. *Revista Electrónica en Actualidades Investigativas en Educación*.
- Picciano, A. G. (2002). Beyond student perceptions: Issues of interaction, presence, and performance in an online course. *Journal of Asynchronous Learning Networks*, 6(1), 21-40. Récupéré du site de la revue : [http://www.aln.org/publications/jaln/v6n1/pdf/v6n1\\_picciano.pdf](http://www.aln.org/publications/jaln/v6n1/pdf/v6n1_picciano.pdf)
- Romero, C., Gutiérrez, S., Freire, M. et Ventura, S. (2008). Mining and visualizing visited trails in Web-based educational systems. Dans R. S. J. de Baker, T. Barnes et J. E. Beck (dir.), *Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings* (p. 182-186). Montréal, Canada : International Working Group on Educational Data Mining. Récupéré du site de la conférence, section Proceedings : [http://www.educationaldatamining.org/EDM2008/uploads/proc/19\\_Romero\\_19.pdf](http://www.educationaldatamining.org/EDM2008/uploads/proc/19_Romero_19.pdf)
- Romero, C., Ventura, S., Espejo, P. G. et Hervás, C. (2008). Data mining algorithms to classify students. Dans R. S. J. de Baker, T. Barnes et J. E. Beck (dir.), *Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings* (p. 8-17). Montréal, Canada : International Working Group on Educational Data Mining. Récupéré du site de la conférence, section Proceedings : [http://www.educationaldatamining.org/EDM2008/uploads/proc/1\\_Romero\\_3.pdf](http://www.educationaldatamining.org/EDM2008/uploads/proc/1_Romero_3.pdf)

- Romero, C., Ventura, S. et Hervás, C. (2005). Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web. Dans R. Ruiz, J. C. Riquelme, J. S. Aguilar-Ruiz et F. J. López Fraguas (dir.), *Actas del III Taller Nacional de Minería de Datos y Aprendizaje* [TAMIDA 2005] (p. 49-56). Récupéré du site de la conférence, section *Libro de actas* : <http://www.lsi.us.es/redmidas/CEDI/papers>
- Rozé, G. (2005). *Structuration et analyse de traces hybrides issues de situations d'apprentissage*. Rapport de Master 2<sup>e</sup> année non publié, Université Joseph Fournier, Grenoble, France. Récupéré du site du directeur de recherche à l'Institut d'informatique et mathématiques appliquées de Grenoble, section *Recherche – Encadrement* : <http://www-clips.imag.fr/arcade/User/jean-philippe.pernin/recherche/encadrement.html>
- Settoui, L. S. (2006, mai). *Systèmes à base de trace pour l'apprentissage humain*. Communication présentée aux Premières Rencontres Jeunes Chercheurs en EIAH, RJC-EIAH, Every, France. Récupéré du site des rencontres, section *articles* : [http://events.it-sudparis.eu/rjc\\_eiah2006/proceeding](http://events.it-sudparis.eu/rjc_eiah2006/proceeding)
- Vandebrouck, F. et Cazes, C. (2004, octobre). *Exploitation des journaux de traces (log) en mathématique à l'université* [version électronique]. Communication présentée au colloque Technologies de l'information et de la connaissance dans l'enseignement supérieur et l'industrie (TICE 04), Compiègne, France. Récupéré du site Technology Enhanced Learning in Mathematics, section *Papers* : <http://www.itd.cnr.it/telma/papers.php>
- Zemsky, R. et Massy, W. F. (2004). *Thwarted innovation. What happened to e-learning and why*. West Chester, PA : Learning Alliance for Higher Education. Récupéré du site de l'organisme, section *Thwarted Innovation* : <http://www.thelearningalliance.info/WeatherStation.html>

## Notes

- <sup>1</sup> Niveau correspondant à peu près au DEUG pour la France et au baccalauréat spécialisé pour le Québec.
- <sup>2</sup> Les codes définis ici n'ayant qu'une valeur illustrative, ils ne sont donc ni présentés ni discutés.