

Canadian Journal of Learning and Technology La Revue canadienne de l'apprentissage et de la technologie

Volume 39(4)

Fall/automne 2013

Learning Analytics: Readiness and Rewards

L'analyse de l'apprentissage : état de préparation et récompenses

Norm Friesen, Boise State University

Abstract

This position paper introduces the relatively new field of learning analytics, first by considering the relevant meanings of both "learning" and "analytics," and then by looking at two main levels at which learning analytics can be or has been implemented in educational organizations. Although integrated turnkey systems or modules are not yet available for review, specific technologies for analyzing online student activities have been implemented and piloted. As a result, this paper concludes by advising an incremental approach to institutional preparedness.

Résumé

Ce rapport présente le domaine relativement nouveau de l'analyse de l'apprentissage en considérant d'abord la pertinence des deux termes "apprentissage" et "analyse", puis en examinant les deux principaux niveaux auxquels l'analyse de l'apprentissage peut être ou a été mise en œuvre dans les organisations éducatives. Bien que des systèmes ou des modules « clés en main » intégrés ne puissent pas encore être examinés, des technologies spécifiques pour analyser les activités en ligne des étudiants ont été mises en place et testées. En conséquence, cet essai conseille en conclusion une approche progressive de préparation institutionnelle.

Introduction: Clarifying and Differentiating Data Sources and their Use in Analytics

Although various definitions of learning analytics have been offered and compared (e.g. see: van Barneveld, Arnold, & Campbell, p. 2012) the following formulation serves as a popular starting point: "learning analytics' refers to 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" (LAK 2011, n.p.). There are a number of key terms or phrases in this definition that would benefit from clarification terms or phrases include "data about learners" and "optimizing and understanding learning." This paper first clarifies these, and in so doing, outlines a functional framework for explaining and understanding various purposes and products related to learning analytics. It then presents an overview of two distinct levels of learning analytics implementations (on the course level and institution-wide), and reviews

applications that are currently available or anticipated for these levels. Finally, this paper recommends and outlines a gradual, phased implementation of learning analytics as appropriate for small to mid-sized online educational institutions.

The 2012 NMC Horizon Report states that learning analytics, as an innovation new to education, presents a "Time-to-Adoption Horizon" of "Two to Three Years." The report adds "learning analytics responds to calls for accountability on campuses and aims to leverage the vast amount of data produced by students in academic activities" (New Media Consortium, 2012, p. 22). Despite its relative urgency, it is widely acknowledged that activity in this field has only just begun. The Wikipedia entry for learning analytics states "much of the software that is currently used [in this field] ...duplicates functionality of web analytics software, but applies it to learner interactions with content" (Learning Analytics, 2013, para. 5). In other words, this software adapts for new, educational purposes technology that has been available for some 15 to 20 years. As a result, it seems appropriate to begin by explaining the most basic aspects of learning analytics, both in terms of the kinds of data to be used in this analysis and the purposes to which this activity is directed.

"Data about learners," as referenced in the above definition involves a number of types of data. These data are generally thought of as being a record of students' key actions in interaction with learning management systems such as Blackboard and Moodle, such as logging on, posting and viewing messages, accessing materials, etc. These data can also include more detailed information generated in these contexts, for example, the analysis of the content of students' postings or other writing, their choices and progress through a specific interactive unit or assignment, or their particular preferences and habits as manifest over a range of tasks and interactions. However, the data used to "understand and optimize" learning and learning environments can also be much more general in nature, obtained from outside of a particular course, learning management system or learning environment. It can include, for example, a record of a student's performance in previous courses, with his or her grade point averages and other information that might be a part of a student's institutional transcript. (The use of students' transcripts and other records of previous performance is commonplace today for admissions requirements to graduate and undergraduate programs, but such records are not typically referenced regularly within individual course contexts.) These data may also include students' responses to inventories (such as Meyers-Briggs tests) or other types of data (e.g., student demographic information). Together, this information about past academic performance, about personal preferences and individual backgrounds is sometimes referred to as "dispositional" data.

The phrase "understanding and optimizing learning" from the definition, above, can similarly be taken to refer to a wide range of possible approaches, analyses and interventions. In the case of many of these, however, the use the term "learning" (appearing both in the definition and in "*learning* analytics" itself) is somewhat misleading. A great deal of data derived from students' activities online (whether it is about course accesses per week, completion of assignments or preferences for text, audio or video), are generally not data about *learning* itself. These data are instead records of mouse-clicks, times and durations of access and navigation paths. They may also include other patterns, but whatever interactions they may be, they are not directly reflective of the *quality* of students' actual reading and study, and even less about their "learning." Simply put, accessing a course or a resource is not the same as gaining in-depth knowledge of its contents, or as effectively working to the attainment of its objectives. The path leading from a

record of online actions to the quality of student engagement or the level of their learning involves a series of inferences. And none of these inferences can be made with absolute certainty. For example, a metric for a given student's "time-on-task" is, with current technology, is more accurately an inference based on a measure of time spent with a given page open in a browser, rather than a direct measure of time devoted to actually studying that page. If the same student was to print or save the contents of that page for study offline, a measure of time-on-task would be further complicated.

However, these kinds of interpretations or guesses can be given greater weight when they are connected with patterns, indications or (even better) with correlations established through research and analysis. The patterns and correlations currently established in this area are quite limited, since they rely on data and student performance generated previously in other or similar technical and academic contexts, and the collection of such data has only just begun (e.g., see: Verbert et al. 2012). A rather small number of relatively clear correlations *have* been established linking some types of student data and student performance in courses. These correlations apply not as much to data from online course activity, as to data from students' transcripts. The best predictor of a given student's future academic performance, as one article points out, is his or her "prior and early performance" (Brown 2012, p. 5). A practical example of a perceived or likely correlation is provided by practice at one institution for regarding a student as being "at risk": The system takes into account both the current course grade and the student's level of online activity. If "grades are high but a student's participation is low, the software ignores the concern, sending out a caution only when both are low," as a relatively early report in learning analytics explains (EDUCAUSE 2011, n.p.). As a further example, quoting Gibbs (2010) and Chikering and others, McFadven and Dawson explain:

Within the teaching context... the most significant predictors of educational gain 'concern a small range of fairly well understood pedagogical practices that engender student engagement' (p. 5). At least a decade of research and writing has demonstrated that learning technologies, when used appropriately, can help educators adopt the 'seven principles of good practice in undergraduate education' and improve the overall quality of an institution's educational delivery...'" (p. 150).

These seven principles of good practice famously identified by Chickering and Gamson include the encouragement of "contact between students and faculty," the provision of "prompt feedback" and the "communication of high expectations" (1987). Of course, there are surely other indicators, indexes, patterns and responses that have yet be discovered or established through practice and analysis.

The various types of data that can be collected, as well as the general kinds of responses, assessments and interventions that can be taken on the basis of these data, are all integrated in figure 1, below.

Figures 2 and 3 are reproductions of figure 1 at a reduced scale. The paths and components that are shown in bold in Figures 2 and 3 indicate how types of data and particular stakeholders or institutional personnel would typically be combined and interconnected in the two "levels" at which learning analytics has been implemented: on the course level (Figure 2) and on the institutional level (Figure 3).



Figure 1: Overview of current and possible future analytics functions and roles.



Figure 2: Analytics components and channels (from Figure 1) used to provide assessment data for instructors and feedback to students, corresponding to part 2, below.



As indicated in Figure 1, above, data generated by the student, left, are of three kinds:

- 1) Transcript data indicating previous performance.
- 2) Inventories of preferences, career orientations, etc.
- 3) Online course activity.

These data are normalized, amalgamated, compared and otherwise evaluated by an analytics system. They are then represented through various methods of simplification, reporting and data visualization, to be used by a number of different stakeholders, including:

1) Instructional personnel

- 2) Support staff
- 3) Administrative personnel
- 4) Students themselves

The way that these data are simplified and represented visually for these individuals is a matter of some interest and speculation in learning analytics. Students, for example, are generally seen as being presented with a "stop light" or "dashboard" reporting system in which their status in the course is communicated directly to them through a green, yellow or red light, or is otherwise quantified and metered as the semester progresses. The student might also receive an automated or semi-automated email sent on the basis of data reviewed by the instructor. The timeliness of such feedback (sent during the semester rather than at its end) as well as its relatively userfriendly nature reflect best practices suggested for such evaluative communication (e.g., Wiggins, 2012). As one would expect, instructional and administrative personnel would typically be offered aggregations of data for groups of students (classes, cohorts, programs), and over longer periods of time (e.g. mid-semester or end-of-semester reports), and they would also be presented with analytics data of greater complexity and variety. The role of the "data wrangler," also shown in Figure 1, is a possible prerequisite for the use of analytics on a broader program or institutional level: "A Data Wrangler," as a JISC report explains, "experiment[s] with different tools to interpret, visualize and share information with academics as a basis for gaining actionable insights"(JISC, p. 5).

These different types of feedback, the data on which they would be based, as well as the type of user to which they would apply, can be indicated by highlighting different elements of the diagram above, as done in Figures 2 and 3. An exploration of these two basic combinations of these elements forms the basis for the next two sections in this paper. The first section looks at feedback on the level of a single course for both student and instructor, and at the existing software that performs this analysis. The second section considers more general feedback suited to administrative purposes, and notes that applications for such purposes are only now being introduced to the market. A third section deals with possible future applications.

Assessments for Instructors and Feedback to Students

As mentioned above, the *provision of feedback to the student* appears to be the most widelydiscussed scenario for the use of analytics in the literature. It also forms the basis for one of the most prominent learning analytics software products currently available, Purdue University's "Course Signals." This analytics package has been under development since 2005, and is designed to be integrated with the Blackboard Learning Management System (LMS). As the name suggests, Course Signals shows students a green, amber or red "traffic" light at various points in the course; and corresponding with this signal, the instructor can also compose and automatically distribute more detailed feedback and tips for improvement to students via email. All of these types of feedback are to be based on a set of calculations customized by the instructor, using a mixture of data from students' past performance, current course grades, and indicative of student "time on task" (based on course and course-tool accesses over a given period; see: Course Signals, 2013a).

Home Course Progress Progress History Settings	Home
Grade & Effort 2 Gradebook 3 Results 4 Review Messages 6 Finish	Grade 8
Section Choose Section	
Grade Criteria	Grade C
Minimum % Maximum %	Grade Gr
Low Risk 100	L/
Medium Risk	Medir
High Risk 0	Hi
Effort Criteria	Effort Cr
A student effort is based on their use of the course website: what portions they use and how frequently they visit it in a given time period. Enter that time period below:	
From To	From

Figure 4: Options for setting grade and "effort" (or time on task) for Purdue's Course Signals application.

These forms of feedback are updated only at intervals set by the instructor over the course of the semester. Instructors are encouraged to update the status of these signals early and often in a given course. Teaching personnel are also advised to provide email messages tailored to the status of particular student groups, and written in the instructor's own voice: "students should see [these as] ...message[s] from you, not from the Course Signals program," as the documentation suggests. Indeed, based on further documentation and tips available for this tool, it is clear that the effective implementation of this software requires significant course instructor (or designer) planning, intervention and integration of its functions with current course designs (e.g., see Course Signals, 2011). At the same time, the types of instructional design features that would be introduced through the careful implementation of this tool – early and detailed feedback based on evaluation throughout the semester – are generally considered to be highly effective for instruction and course design overall. Findings highlighted on the tool website seem to reflect this alignment between the design of Course Signals and good instructional practice – although these findings appear rather anecdotal in character (see Course Signals, 2013b, n.p.).

A second specialized analytics tool, which has similarly been under development for a number of years is SNAPP, short for "Social Networks Adapting Pedagogical Practice." This tool analyzes one particular kind of online course data: student participation in online discussion forums. It is designed to provide summaries of this participation for the benefit instructional and advisory personnel, rather than offering ways of directing customized feedback to students. Asynchronous text-based discussion, of course, has long been a mainstay of distance and blended learning (e.g., Garrison & Vaughan, 2008), and SNAPP provides ways of gauging and visualizing student activity in these fora – specifically for Blackboard, Moodle, Desire2Learn and Sakai– in the form of a social network graph (<u>http://www.snappvis.org/</u>). Here is how the product documentation describes its functions:

SNAPP infers relationship ties from the post-reply data and renders a social network diagram below the forum thread. The social network visualization can be filtered based upon user activity and social network data can be exported for further analysis in NetDraw [free software for rendering network data]. (SNAPP, 2013, n.p.)

SNAPP works without installation or tweaking simply as a browser "bookmarklet." It renders diagrams that show the degree to which students are active and/or central in exchanges of messages and replies, and that help to identify students and patterns of discussion isolated from the class, as well as interactions that are more or less independent from instructor orchestration.



Figure 5: SNAPP – Social network graph of 177 posts contributed by six participants

The way that social network analysis (as utilized in SNAPP) can be interpreted by the instructor is explained in the user guide: Each node represents a participant; the size of the node is in proportion to the number of messages sent (by default, but evidently not the diagram above); the quantity of messages sent and received are indicated on the lines between the nodes. The patterns of nodes and lines can be seen to indicate both effective and less desirable patterns of communication, thus providing a kind of feedback that in this sense goes beyond the strictly quantitative forms to which many analytics applications are limited. However, this application is similar to many other analytics applications in that it does not reflect the quality of student participation *directly*, but only through inference. Thus some nodes or participants who may appear very active or communicative may actually only be sending and receiving very brief, superficial communications (e.g., "awesome post!"), rather than the longer or more reflective messages that may be more indicative of deep processing or active learning.

Analytics for Program- and Institutional-Level Decision-Making

A recent report from the US Department of Education makes the point that on the program and institutional level, learning analytics can play a role that is similar to that of already existing business intelligence departments and applications:

Using data for making decisions is not new; companies use complex computations on customer data for *business intelligence* or *analytics*. Business intelligence techniques can discern historical patterns and trends from data and can create models that predict future trends and patterns. Analytics, broadly defined, comprises applied techniques from

computer science, mathematics, and statistics for extracting *usable information* from very large datasets. (Bienkowski, Feng & Means, 2012, p. 7)

Just as business intelligence may utilize demographic, behavioral and other information associated with a particular enterprise and its customers to inform decisions about marketing, service and strategy, learning analytics promises to do something similar in educational terms. Students might well be better served through processes of admissions, support and instruction that are directly responsive to common patterns and needs evident in related academic contexts. This may be particularly the case if performance, attrition and completion are not simply collected, but also are able to be seen at a glance, over different periods of time, institution-wide. Also, coordinating, cross-referencing and visualizing this data in different ways may provide new insights on "pressure points" for student satisfaction, performance and retention. Indeed, for those educational organizations already utilizing business intelligence methods for planning purposes, learning analytics should be able to augment and also make use of information already available for planning support and deployment. However, as mentioned at the outset of this paper, the types of services and systems for achieving this high-level analysis are currently only now emerging. Documentation of feature sets, independent reviews of capability and usability, and other information for such products and services - beyond vague marketing claims and promises - are as of yet unavailable.

At the same time, an informative example of the *possibilities* offered by these "high-level" learning analytics applications is provided in a recent article by Macfadyen & Dawson (2012). It reports on the results of a custom analytics application developed using WebCT Vista at the University of British Columbia (UBC). The analytics software was provided by MicroStrategy Business Intelligence software, and data visualization was made possible by Tableau Software. Both systems were customized for the analysis of aggregate use of LMS services at the university, together with the analysis of student grades and individual student behavior online, as the article explains:

To overcome the challenge of poor analytics functionality in the current LMS, the university considered here has partnered with an analytics software company to customize and implement an analytics reporting tool that allows extraction, analysis and dis/aggregation of detailed information about uptake and use of the enterprise LMS. We made use of this analytics platform to carry out the requested "current state analysis" of LMS usage, and to seek answers to questions about the extent and complexity of LMS adoption. This analysis was undertaken with the goal of informing and guiding the institution's campus-wide strategic planning process for learning technology and LMS integration. (Macfadyen & Dawson, 2012, p. 151)

The richly documented results of this study include information on user time online in the LMS, as well as the particular LMS components and types of activity (administration, assessment, viewing and engaging with content) across which this user time was reported to have been distributed. The study's findings include the identification of a "significant correlation between student achievement and selected LMS tool use frequency in LMS-supported course sections":

Correlation coefficients for the selected LMS activities shown here with binned student final grade are as follows: number of discussion messages posted, r = .83, p<.01; number of discussion messages read, r = .95, p<.0001, number of discussion replies posted, r = .94,

p<.0001); number of content pages viewed (0.89, p<.001); number of visits to the "My Grades" tool (0.93, p<.0001). (2012, p. 157)

To simplify, active student participation in online discussion and student self-monitoring of performance during a given course, showed the strongest correlation with student achievement in that course. This is a statistical relation that is significant, but that does not indicate causality.

It is important to note that Macfadyen and Dawson report title their report as follows: "Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan." Combined with the authors' final conclusions, this title highlights a key component that Macfadyen and Dawson believe was missing in their introduction of analytics to UBC: This is a sense of the meaning and potential of learning analytics in the larger organization itself, and shared frameworks for understanding the results produced by analytics, and a willingness to use them in policy development. As Macfadyen and Dawson emphasize, the nascent state of the analytics field has rendered "analytics data difficult for non-specialists to generate (and generate in meaningful context), to visualize in compelling ways, or to understand, limiting their observability and decreasing their impact" (2012, p. 161). This highlights the importance of the further development and integration of learning analytics systems with existing institutional metrics and reporting mechanisms – as well as further explanation and demonstration of the value of such systems.

Learning Analytics in the Future

As noted above, learning analytics both "responds to calls for accountability" and is also able to leverage vast amounts of data now produced through students' academic activities. Given these two powerful, positive factors, it is very likely that the learning analytics field and the technologies associated with it will be developing further and maturing in the short to middle term. This may mean that some of the challenges posed to non-specialists as noted by Macfadyen and Dawson, above, may be addressed relatively quickly. Also, given the increasing importance of efforts related to student engagement and experience in higher education, and to the scholarship of teaching and learning in many universities, recognition of the value of learning analytics as a means of tracking and reporting on student participation and activity may not be lacking for long. Ongoing ICT integration in university environments only makes such developments all the more likely.

However, there are also factors that I believe may have the opposite effect, *diminishing* the potential recognition and maturation of learning analytics as an accepted part of instructional practice and academic administration. The most significant of these, I believe, is its conceptualization and promotion as offering a kind of educational "silver bullet," as a solution for a range of perennial educational challenges, and as a means of enforcing comprehensive regimes of assessment and administration. For example, the US Department of Education report mentioned above predicts that with sophisticated analytics,

online or adaptive learning systems will be able to exploit detailed learner activity data not only to recommend what the next learning activity for a particular student should be, but also to predict how that student will perform with future learning content, including high-stakes examinations. (2012, p. 3)

Students' needs and their learning preferences are complex and vary over time; something similar can be stated about their engagements with printed text, to say nothing of their interactions with online and/or multimedia content. Such interactions and engagements have been subject to a great deal of research, and are still not comprehensively, or in some cases, even partially understood. Even less understood, as indicated above, are the causal relationships between these kinds of student activities and student academic performance. Finally, earlier experiments and efforts to implement adaptive computer-based instructional systems into institutional instructional environments have generally not met with success (e.g., Gibbons & Fairweathe, 1998, pp. 280-282). It is therefore rather risky to assume that learning analytics may both clarify these understandings, and automate appropriate responses and interventions for students on this basis. These types of assumptions reiterate promises about the "automation" of instruction that have been made repeatedly in the past and that have also repeatedly disappointed. This includes, for example, predictions made in the contexts of the programmed instruction and learning objects movements (both saw content as being customized according to students' previous interactions). Adaptive systems not only work against the widely-promoted benefits of learner self-monitoring and self-regulation (e.g., Zimmerman & Schunk, 2011), they also undermine the goal of students becoming autonomous and life-long learners. And to associate learning analytics with controversial regimes of high-stakes testing, finally, would mean to politicize it and cast it into doubt (through association) in the minds of many practitioners. In short, to focus on the incremental improvements in instruction and administration that learning analytics is indeed likely to attain is more beneficial than promoting advancements that it most likely will not.

At a time when institutional accountability is becoming increasingly important, and quantities of student data are growing ever larger, the potential of learning analytics as an adjunct to established institutional metrics is difficult to dispute. However, in order to realize this potential, it is just as important not to oversell or exaggerate these benefits, as it is to explain the very practical gains that they might represent.

References

Bienkowski, M., Fieng, M. & Means, B. (2012). enhancing teaching and learning through educational data mining and learning analytics. U.S. Department of Education Office of Educational Technology. Retrieved from http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf

Brown, M. (2012). Learning Analytics: Moving from Concept to Practice. *EDUCAUSE Learning Initiative*.

Chickering, A. W. & Gamson, Z. F. (1987). "Seven principles for good practice in undergraduate education" *American Association of Higher Education Bulletin*, 39(7) 3-7.

Course Signals. (2011). Top 10 tips. Retrieved from: http://www.itap.purdue.edu/learning/tools/signals/Top Ten Tips.pdf

- Couse Signals. (2013). How does Course Signals work? *YouTube Video*. <u>http://www.youtube.com/watch?feature=player_embedded&v=KxYewO9iAgw#action=sh</u> <u>are</u>
- EDUCAUSE (2011). 7 Things you should know about first generation learning analytics. *Analytics Series*, 1(8). Retrieved from <u>http://www.educause.edu/library/resources/7-things-</u> you-should-know-about-first-generation-learning-analytics
- Garrison, D. R., & Vaughan, N. (2008) Blended Learning in Higher Education: Framework, Principles, and Guidelines. San Francisco: Jossey-Bass.
- Gibbons, A. S. & Fairweather, P. G. (1998). *Computer based instruction: Design and development*. Englewood Cliffs NJ: Educational Technology Publications.
- Gibbs, G. (2010). Dimensions of quality. York, UK: The Higher Education Academy. Retrieved from <u>http://www.heacademy.ac.uk/resources/detail/evidence_informed_practice/Dimensions_of_Quality</u>
- LAK 2011. (2011, February March). 1st International Conference on Learning Analytics and Knowledge. Banff, Alberta, February 27–March 1, 2011.
- Learning analytics. (n.d.). In *Wikipedia, The Free Encyclopedia*. Retrieved April 18, 2013, from http://en.wikipedia.org/w/index.php?title=Learning analytics&oldid=550997848
- Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15(3), 149–163.
- New Media Consortium. (2012). The NMC horizon report: 2012 higher education edition. Retrieved from <u>http://www.nmc.org/publications/horizon-report-2012-higher-ed-edition</u>
- Noë, A. (2010). No out of our heads: Why you are not your brain, and other lessons from the biology of consciousness. New York: Hill and Wang.
- Powell, S. & MacNeill S. (2013). Institutional readiness for analytics. *Analytics Series*, 1(8). Retrieved from <u>http://publications.cetis.ac.uk/wp-content/uploads/2012/12/Institutional-Readiness-for-Analytics-Vol1-No8.pdf</u>
- SNAPP. (2013). *About SNAPP: Social Networks Adapting Pedagogical Practice*. Retrieved from http://www.snappvis.org/?page_id=4
- van Barneveld, A., Arnold, K. E. & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative*. Retrieved from: <u>http://net.educause.edu/ir/library/pdf/ELI3026.pdf</u>
- Verbert, K., Manouselis, N., Drachsler, H., Duval, E (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society 15*(3), 133-148.

Wiggins, G. (2012). "Feedback for Learning: Seven Keys to Effective Feedback." ASCD Educational Leadership, 70(1), 10-16.

Author

Norm Friesen. Email: normfriesen@boisestate.edu.



This work is licensed under a Creative Commons Attribution 3.0 License.