Learning Analytics: The Coming Third Wave

By Malcolm Brown, Director, EDUCAUSE Learning Initiative

April 2011

» Learning analytics (LA) is the third wave of large-scale developments in instructional technology that began with the advent of the learning management system.

» As a compelling diagnostic tool, LA will see rapid adoption over the next several years, as third-party applications begin to make it more affordable and practical.

» The papers presented at the First International Conference on Learning Analytics and Knowledge (LAK11) conference provide an overview of current research and development, helping to show the potential of LA as a way to provide timely and meaningful feedback for learners, instructors, and administrators.

Learning analytics is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections for predicting and advising people's learning.
—Wikipedia

Learning analytics is the use of data and models to predict student progress and performance, and the ability to act on that information.
—Next Generation Learning Challenges

What Is Learning Analytics?

At its core, learning analytics (LA) is the collection and analysis of usage data associated with student learning. The purpose of LA is to observe and understand learning behaviors in order to enable appropriate interventions. The reports that an LA application generates can be very helpful for instructors (about student activities and progress), for students (feedback on their progress), and for administrators (e.g., aggregations of course and degree completion data). The main elements of LA include:

■ **Data collection:** This entails the use of programs, scripts, and other methods to gather data. This can be data from a single source or a variety of sources; it can entail large to very large amounts of data, and the data can be structured (e.g., server logs) or unstructured (e.g., discussion forum postings). The specific design of the collection activity is informed by the goals of the LA project.

■ **Analysis:** Unstructured data is usually given some kind of structure prior to analysis. The data is subjected to an appropriate combination of qualitative and quantitative analysis. The results of the analysis are reported using a combination of visualizations, tables, charts, and other kinds of information display.

■ **Student learning:** This core goal distinguishes learning analytics from other kinds of analytics. LA seeks to tell us about student learning: what learners are doing, where they are spending their time, what content they are accessing, the nature of their discourse, how well they are progressing, and so on—at the individual or cohort level or both.

■ **Audience:** The information that LA returns can be used to (1) inform instructors, (2) inform students, or (3) inform administrators. Common to all three is that the reports enable appropriate interventions. Typically (1) and (2) enable course-level interventions, while (3) informs interventions at the departmental, divisional, and institutional levels. The kinds of data and analysis employed depend on the intended audience.

■ **Interventions:** The reason for doing LA is to enable appropriate interventions at the individual, course, department, or institutional level. LA can do more than just identify students at risk. By analyzing the digital

"breadcrumbs" that students generate when participating in all aspects of a course, it is possible to observe student progress at specific stages and at specific activities in a course. The potential of LA is to be able to indicate what is working and what is not at a much finer level of granularity than ever before, even while a course is in progress.

Adoption Horizons

The *2011 Horizon Report* includes LA in the four- to five-year window, meaning it predicts LA will be in "mainstream use" within that time. This estimate, however, might too conservative. A combination of factors makes a two- to three-year horizon quite probable.

> *LA will provide the capability of collecting and analyzing data from a variety of sources to provide information on what works and what does not*

The first factor is that LA is very compelling. The way that LA reveals usage, trends, and patterns in learning is by itself compelling enough to spur adoption. In addition, LA helps identify students who are struggling, enabling interventions that can assist those students to successful course completion. This is of particular importance, given the enormous pressures on higher education for accountability and higher completion rates. LA is perhaps the key resource that will enable schools to begin to respond to those calls.

Traditional student course evaluations capture student opinion and reflections but cannot capture what the students actually did. Moreover, because those evaluations are done once the course is complete, they cannot enable interventions while the course is in progress. Nor can they provide any information on things such as how students utilized course content or the extent of their participation in course activities. LA can do that and more.

The second factor is the emergence of LA applications. Until recently, if a school wanted to conduct LA, it had to build a system for LA from scratch. Two examples of this include Purdue University's Signals application and the development work done at the University of Maryland, Baltimore County, to coax LA from its Blackboard implementation. This landscape is now changing rapidly. SunGard, Blackboard, and Desire2Learn are moving in this direction, and over the next year, more vendors will enter this space. This development is important in that it enables schools to begin to do LA without costly and time-consuming development. In

short, we are now entering a time when institutions can buy instead of build and can have an LA application up and running in a few months instead of years. This factor, even more than the others, might accelerate the adoption of LA at institutions worldwide.

A third factor is the increasing emphasis on the use of metrics and rubrics for higher education in order to demonstrate student learning and degree progress. This will encourage the investigation of LA as a way of meeting those demands. In addition, foundations are now placing a heavy emphasis on the use of LA in the projects they support. A example is the Next Generation Learning Challenges program, which has issued a challenge to "develop a model that identifies, improves, and scales existing solutions of learner analytics."[1]

In suggesting a more rapid adoption of LA, we are not claiming that within three years the majority of schools will be doing a sophisticated, highly customized form of LA. That will take more time. But given the environmental factors mentioned above, it is clear that the majority of schools will be "doing" LA in some form, even if it is simply installing an off-the-shelf application.

LA as the Third Wave

LA is the third wave of developments in instructional technology that began with the advent of the learning management system (LMS). In the late 1990s it was clear that handcrafting course websites would not scale, and so the LMS was born. By 2005, the question for schools was no longer whether they were running a LMS but instead which one. The vision underlying the initial LMS design was based on a traditional teaching model, being course- and instructor-centric, and the transmission paradigm of Web 1.0. Nevertheless, the LMS has seen large adoption rates, as its utility is clear. The adoption of the LMS 1.0 also enabled a key accomplishment: integration into the enterprise. Now learning, for the first time, was connected to the enterprise infrastructure.

The second wave is the addition of a "2.0" layer to the LMS, adding social networking and cloud-based applications into the mix. At times this has been the proverbial square peg in a round hole, given the "1.0" vision that informed much of the LMS initial design. Nevertheless, because the 2.0 functionality has been so compelling, higher education has moved, and rapidly, in this direction. Indeed, this has been so important that faculty and others have at times gone outside the LMS in search of 2.0 functionality. Today we are rapidly approaching the point at which the question is not whether the 2.0 tools are being used but rather which ones and in what manner.

LA will be the third wave. Building on the LMS and its enterprise integration, LA will provide the capability of collecting and analyzing data from a variety of sources to provide information on what works and what does not with respect to teaching and learning. LA's strength will come from the enterprise integration that the LMS established, and it will become more compelling as data retrieval and analysis become more powerful and sophisticated. The new diversity of 2.0 tools will offer a larger set of digital "breadcrumbs" and hence more data. Colleges and universities might even be able to share anonymous, aggregated data to produce composite views of student learning for an entire class of institutions.

The LMS addressed the administrative overhead of running a course, and the 2.0 wave is producing learner-centered content and tools. LA, as the third wave, is the metacognitive component, allowing individuals and institutions to understand learning and make informed decisions about resource allocations and required interventions to promote learner success.

Privacy: The Ethics of Doing LA

At the first International Conference on Learning Analytics and Knowledge (LAK11), held in Banff, Canada, February 28–March 1, 2011, presenters and attendees all agreed that LA raises deep and complex privacy issues. LA could be construed as eavesdropping. The ethical concerns are obvious. The worry is that the LA mechanisms will prowl about, scrutinizing what learners are doing, and so constitute an invasion of privacy.

One obvious way to address privacy concerns is to allow students to opt-out. But there are problems with this approach. The accuracy of LA's recommendations for intervention improves as (1) the number of observable "subjects" increases and (2) the amount of data available for analysis is maximized. Analysis based on only a subset of the students in a course or based on fragmentary data will be incomplete. That could mean that the resulting recommendations for intervention might be less accurate. As a result, an opt-out alternative for students could compromise the LA effort. Much of the future work around LA will involve finding the balance between learner privacy and the value of data collection for improving learning.

One presenter at the conference, Eric Duval, touched on this issue. He suggested the idea that the person who generates the breadcrumb data owns that data and that the data should be "refundable" or returnable to the owner. He mentioned a project, Attention Trust http://www.attentiontrust.org/, that for a time was developing this idea of refundability. Yet this idea would seem to return us to the problem of accuracy, for data refunds would create holes in the data, which could in turn lead to incomplete analyses.

Clearly a good deal of work needs to be done here. As schools proceed, it will be necessary to engage a variety of campus stakeholders to ensure a balance between the demands of individual privacy and the goals of a LA project.[2]

Selected Presentations from the LAK11 Conference

The presentations at LAK11 presented a compelling picture of the potential of LA. We summarize just a few of the papers to try to capture a sense of this.

- A presentation by Ravi Vatrapu (Copenhagen Business School) introduced the idea of "teaching analytics" and reported on a project designed to provide the instructor with real-time analytics information while in the classroom. The goal is to "provide timely, meaningful, and actionable formative assessments to ongoing learning activities in situ." Their system consists of three computer "experts"; these three experts work by "analyzing, interpreting and acting upon real-time data being generated by students' learning activities by using a range of visual analytics tools."

- An interesting presentation by Chris Brooks (University of Saskatchewan) reported on using a lecture-capture system to do LA. Utilizing a presentation-capture system designed in-house, they have designed the tool to report data on the use of lecture-capture content by the students. The usage information sent back to the server is detailed enough to largely reconstruct the student's session with the content. The goal of the project is to understand how students use this system to augment learning and to move toward validating the constructivist educational theory. The information harvested by their system has "been used to help analyze the behavior of hundreds of students over an academic term, quantifying both the learning approaches of students and their perceptions on learning with lecture capture."

- Simon Buckingham Shum (Open University) reported on research using LA to identify exploratory dialogue (following a discourse typology framework by Mercer and colleagues). In a learning context, exploratory dialogue represents a high-order level of discourse because it entails reasoning, challenge, and evaluation, all three of which are markers or indicators of deeper learning. The other two kinds of discourse, disputational and cumulative, involve fewer of those three elements. The challenge is to design an LA tool that can identify exploratory dialogue,

flag it for the instructor and student, and locate "dialogue exchanges where deeper learning appears to be taking place." This is research in progress; they are seeking to confirm whether their initial set of exploratory dialogue markers really do work as identifiers.

- Mike Sharkey (University of Phoenix) reported on their LA efforts. His presentation conveyed the scale at which LA can be conducted. The University of Phoenix data repository draws from over 30 sources, is modeled with 430 tables, and is 1.5TB in size (growing by 100GB per month). The goal of their project is "to predict a student's persistence in their program using available data indicators such as schedule, grades, content usage, and demographics."

- Katja Niemann from Bonn, Germany, presented a way to use LA to assist in content discovery in learning. Learners must make decisions about what to do next and what learning objects (LOs) to use next. The challenge is how then to make suitable recommendations for LOs, especially for resources the learners may not know about—in other words, how to make semantic linkages apart from usage metadata. Their work is based on the linguistic theory of co-occurrence, which contends that the closer words appear in discourse, the more semantically relevant they are. They have performed initial analyses of 4,000 objects and have statistically relevant results.

- Aneesha Bakharia (University of Queensland) and Shane Dawson (University of British Columbia) presented on SNAPP (Social Networks Adapting Pedagogical Practice). This tool, by making available "data visualisations and social network metrics in 'real-time', allows emergent interaction patterns to be analyzed and interventions to be undertaken as required." Hence, "SNAPP essentially serves as an interaction diagnostic tool." The presentation described the features that will be added in version 2: "The ability to view the evolution of participant interaction over time and annotate key events that occur along this timeline. This feature is useful…[for] evaluating the impact of intervention strategies on student engagement and connectivity."

To Explore Further

- Website for the First International Conference on Learning Analytics and Knowledge 2011, https://tekri.athabascau.ca/analytics/.
- George Siemens, "Learning Analytics: A Foundation for Informed Change in Higher Education," ELI Webinar, January 2011, http://www.educause.edu/Resources/LearningAnalyticsAFoundationfo/221468.
- David Wiley, "Openness, Learning Analytics, and Continuous Quality Improvement," podcast, http://www.educause.edu/blog/gbayne/ELI2011PodcastOpennessLearning/225360.
- *2011 Horizon Report*, http://www.educause.edu/Resources/2011HorizonReport/223122.
- International Working Group on Educational Data Mining, http://www.educationaldatamining.org/.
- John Fritz, "Video Demo of UMBC's 'Check My Activity' Tool for Students," *EDUCAUSE Quarterly* 33, no. 4, 2010, http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/VideoDemoofUMBCsCheckMyActivit/219113.
- Kimberly E. Arnold, "Signals: Applying Academic Analytics," *EDUCAUSE Quarterly* 33, no. 1, 2010, http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/SignalsApplyingAcademicAnalyti/199385.
- Syllabus to the open online course on Learning and Knowledge Analytics, http://learninganalytics.net/syllabus.html.
- Donald Norris, Linda Baer, Joan Leonard, Louis Pugliese, and Paul Lefrere, "Action Analytics: Measuring and Improving Performance That Matters in Higher Education," *EDUCAUSE Review* 43, no. 1 (January/February 2008): 42–67, http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/ActionAnalyticsMeasuringandImp/162422.
- John P. Campbell and Diana G. Oblinger, "Academic Analytics," EDUCAUSE White Paper, October 2007, http://net.educause.edu/ir/library/pdf/PUB6101.pdf.

Well over two dozen papers were presented at LAK11. Overall, they represent the rich diversity of research and project directions currently under way in the field of LA. One analogy used at the conference is that LA enables instructors, learners, and administrators to "take the temperature" of the learning that is taking place at our institutions. This ability to "look inside" learners' activities provides rich diagnostic information, much the way that an X-ray or an MRI scan provides information for medical diagnosis. This is what makes the case for LA so compelling.

Endnotes

1. See http://nextgenlearning.org/the-grants/wave-1-challenges/learning-analytics/articles/learning-analytics.
2. FERPA does permit disclosure of education records without prior consent to school officials with "legitimate educational interests." The full wording in the statute reads, "The disclosure is to other school officials, including teachers, within the agency or institution whom the agency or institution has determined to have legitimate educational interests." See 34 CFR § 99.31, http://edocket.access.gpo.gov/cfr_2004/julqtr/pdf/34cfr99.31.pdf.