EDUCAUSE LEARNING INITIATIVE BRIEF

Learning Analytics: Moving from Concept to Practice

Malcolm Brown, Director, EDUCAUSE Learning Initiative

July 2012

- **Amid a long list of measurable factors, some have been shown to correlate strongly with academic outcomes, while others are not strong indicators of student success.**
- **The representations—often graphical—of the patterns and insights gleaned from analytics are a central component of how that information is understood and used.**
- **The most effective learning analytics programs will be institution-wide efforts, taking advantage of a wide range of resources and possible interventions.**

Analytics is an umbrella term for the use of data, statistical analysis, and explanatory and predictive models to gain insights and act on complex issues. As a genre of analytics, learning analytics (LA) uses these methods to achieve greater success specifically in student learning. LA can be used in a variety of ways, some of which include alerting faculty, students, and advisors when intervention is needed; providing input for continuous improvement in course design and delivery; and enabling personalization of the learning environment.

In 2011, ELI issued a brief that described LA as the coming third wave, a new technology with great potential to increase student academic success. This second ELI brief on LA describes the evolution of the topic over the past year. It draws on two major recent conferences devoted to the theme of learning analytics: the ELI 2012 Spring Focus Session (SFS), which was held on April 11 and 12, 2012, and the second Learning Analytics and Knowledge conference (LAK12), convened April 29–May 3, 2012.

The structure of this brief follows the major themes that emerged at these events:

- Definitions (what distinguishes LA from other analytics)
- Predictors and indicators (the data that LA uses)
- Visualization (rendering visible and accessible the results of data analysis)
- Interventions (actions and decisions undertaken based on LA results)

Definitions

LAK12 defined learning analytics as "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."[1] An important characteristic of this definition is its distinction between LA technology and the purposes it serves. At the 2012 ELI SFS, George Siemens drew the same distinction, remarking that "all the important stuff with analytics happens…after we've done the analytics."[2] Analytics technology provides information and evidence that enables what he calls "sensemaking" and what elsewhere is called decision making. This distinction is of key importance, as any institutional program involving LA must have both: a robust technology to harness and analyze data and effective plans and processes for acting on the results of the analysis.

Predictors and Indicators

On the technology side, one of the key decisions when designing an LA application is which data to use as predictors and indicators of student progress. In any analytics initiative, the selection of data directly affects the accuracy of the predictions and the validity of the analysis. Almost every presentation at these conferences addressed this issue, and this discussion will clearly occupy researchers and practitioners over the next several years. But even at this relatively early stage in LA's development, some initial patterns are coming into view.

Dispositional Indicators

These are factors or dispositions brought to the learning context by the learner. They are in place before the course begins and provide some indication of how a student is disposed to and prepared for his learning, and they can also point to how a learner is likely to respond to any course-related interventions.

Many of these factors are factual and are readily quantifiable, such as age, gender, ethnicity, current grade point average (GPA), prior learning experience, first-in-college, or even financial status. Some of these are powerful predictors: Several researchers remarked how a student's GPA alone can be an accurate predictor of a student's performance in a course. For example, Tim McKay of the University of Michigan wrote, "While a number of parameters correlate with final grade, prediction with a half letter grade dispersion can be accomplished using just one parameter: each student's University of Michigan GPA at the start of the term."[3]

Several researchers remarked how a student's GPA alone can be an accurate predictor of a student's performance in a course

On the other hand, a few LA projects have moved beyond such information to also include psychological measures of disposition. The paper given by Simon Buckingham Shum and Ruth Crick at LAK12 provided an interesting example of this idea. Their project is focused "the challenge of designing learning analytics that render visible learning dispositions and the transferable competencies associated with skillful learning in diverse contexts."[4]

One such measure is "learning power," which is defined in Wikipedia as "the collection of psychological traits and skills that enable a person to engage effectively with a variety of learning challenges."[5] The goal of the Buckingham Shum/Crick research is to see if there is correlation between a student's learning power profile and their academic success. Students create their profile by filling out a questionnaire called Effective Lifelong Learning Inventory (ELLI), containing about 75 questions. The responses are used to create a spider-diagram profile that measures seven dimensions, including traits such as resiliency, creativity, critical curiosity, and strategic awareness. Their work is investigating whether this measure can be correlated to higher attainment in coursework, such as grades and test scores. At the moment, while there is not enough evidence to justify a conclusion, there are promising indications: "Consistent with this line of thought, one would predict ELLI to correlate positively with conventional attainment analytics, and indeed, several studies do report a positive correlation… this is an intriguing finding, but this relationship requires further interrogation…."[6]

At the ELI SFS, Clint McElroy of Central Piedmont Community College, a Next Generation Learning Challenges (NGLC) grant recipient, reported on their Online Student Profile Learning System. Starting with a course roster, the instructor can request more detailed information about a student. This system uses dispositional information to help paint a more complete portrait of students. Their design includes both Jungian personality type (similar to the Myers-Briggs and Keirsey inventories) and learning style information in the dashboard that is displayed to the instructor. The instructor sees both current course performance and these dispositional indicators, a blend that helps the instructor make a more accurate determination whether an intervention is required and, if so, what kind.[7]

Activity and Performance Indicators

These measures are the digital breadcrumbs left by learners as they engage in their learning activities and make their way through the course sequence. Many of these are quantitative in nature. Most projects are gathering these breadcrumbs where they are most plentiful: via the learning management system (LMS). Examples include the number and frequency of LMS logins, the amount of time spent on the course website, the number of discussion forum posts, grades, and formative quiz scores. These indicators are relatively straightforward to collect and can be readily analyzed and the results displayed in visualizations.

In Purdue University's Signals application, as reported by Kim Arnold and Matthew Pistilli, two of the four factors used in their student success algorithm to determine a signal are activity indicators: "performance, measured by percentage of points earned in course to date; effort, as defined by interaction with…Purdue's LMS, as compared to students' peers." The other two are dispositional: academic preparation (high school GPA and standardized test scores) and student characteristics (residency, age, or credits attempted).[8]

As Mike Sharkey reported, the University of Phoenix has been working on LA for some time now. The university has refined its prediction model and has tested it using a very large data set drawn from student data in past courses. The model attempts to predict if students will pass their current courses, and these tests have produced accuracy rates ranging from 78% to 92%. Although the program remains very much a work in progress, the university has identified a preliminary set of five "good" or key indicators: scores in the current course, credits earned divided by credits attempted, the delta between past and current scores, GPA in prior courses, and financial status. This is again a blend of dispositional and performance indicators. Also of interest are indicators the university has found to have little or no predictive value: date of first course activity, number of concurrent courses, gender, military status, and ethnicity.[9]

By contrast, an early-warning system developed for engineering courses at the University of Michigan relies exclusively on activity indicators. According to Steve Lonn and Stephanie Teasley, this project uses LMS data, including gradebook and assignment tool data. They also "integrated a proxy for student 'effort' in a course through the use of LMS course website login events." Note that this project uses the same proxy for student effort as does the Purdue Signals program. In the Michigan project, these data were aggregated, displayed in a visualization dashboard, and made available to mentors.[10]

The work presented by Al Essa and Hanan Ayad of Desire2Learn uses a blend of five indicator types or indices as the foundation for its predictive modeling and analysis. They call this the Success Index. One index is dispositional (preparation), while the remaining four are activity-related (attendance, participation, completion, and social learning).[11]

What emerged at these conferences is that, at this time, most projects use a blend of dispositional and performance indicators and that these deliver strong predictive capability. Over time, we can expect to see more sophistication with a more heterogeneous data types blended to produce more accurate and nuanced predictions.

Student Artifacts

These are the actual work products of the students: their essays, blog and discussion forum posts, media productions, and so forth. Some contend that direct analysis of such artifacts can provide indications of whether students are achieving the needed level of expertise and whether they are exhibiting higher-order thinking skills in their work. This approach is far less common but has the potential to put LA "closer" to the actual learning by detecting indications of competence and mastery.

The work of Lárusson and White (Brandeis University and UC Berkeley, respectively) provides one example. Their application examines student writing to determine if the student has progressed from mere repetition of what is heard and read to original thought about course content. Using a lexical resource, the tool generates originality scores for student work. Hence their application (called the Point of Originality tool) "seeks to gauge [the] students' ability to interpret, to place core concepts into new and diffuse usages. This definition of originality straddles the tiers of learning that Bloom's taxonomy associates with 'understanding' and 'application.'"[12] Their initial test of the tool, analyzing student blog postings, produced encouraging results, showing "that the tool was generating originality scores for students' blog posts that correlated both with the degree to which they participated in the online activity as well as the final grades that they received for their term papers."

Visualization

Visualization and reporting are the elements that make LA's intelligence truly actionable. These make visible the patterns in the data, enabling instructors, advisors, and students to take appropriate actions. The visualization component of LA has two aspects: the way the results of analysis are displayed (the type of chart or diagram used), and dashboarding (the way visualization elements or widgets are selected and combined for review by instructors, advisors, and students).

> ***[Visualizations] make visible the patterns in the data, enabling instructors, advisors, and students to take appropriate actions***

There seem to be two approaches to dashboards. The work of Santos and Duval is an "all-at-one-time" approach, whereby a variety of visualization windows are displayed side-by-side, each providing a different "view." In their initial design, their dashboard comprises eight visualization windows or widgets, and much of their research is focusing on determining which displays are useful to engineering students and which are not. In preliminary findings, they have found that while students value reports that compare their efforts to others in the class, a few students mentioned that they dislike having others see their activity.[13]

By contrast, some approaches—like the one described by Essa and Ayad—start with a single visualization window and allow the user to build from there. In their project, they start with a summative, traffic light–like display of overall student status and risks. Their system allows the viewer to then drill down and retrieve more detailed information. For example, the viewer can pull up a student profile screen, a course screen that displays the student's course-level activities and risks, a notes screen for case notes, and a referral screen that "provides all the relevant referral options available at the institution."[14]

Across these approaches, it is clear that visualization and user-interface design play a key role: They unlock information and make it compelling and accessible to faculty, students, and advisors.

Interventions and Responses

If LA technology aims to produce actionable intelligence, intelligent actions are required. Lonn and Teasley introduced the term *organizational capacity* as a label that captures the action side of the LA coin and defined it as "the resources and routines that both enable and constrain access and application of learning analytics tools and related services." To be truly effective, a program to promote learning success built around LA needs to be institutional in scope, marshaling a variety of institutional resources, especially with respect to interventions. Given this, it would seem to make sense to recast this term as *institutional capacity*. Institutional capacity is the sum of the resources needed to build, run, and maintain the LA technology and the resources to act effectively on LA's intelligence.[15]

Designing effective interventions is no small task

Designing effective interventions is no small task. At a recent ELI webinar, Kay McClenney, director of the Center for Community College Student Engagement (CCCSE), shared evidence that student orientation programs and academic success courses—both forms of intervention—can contribute substantially to student success. In order to realize this potential, however, these resources must be carefully crafted so as to be relevant to students' actual needs. Her study revealed that if students perceive the resources as rote, impersonal, and superficial, they pay little or no attention. It is not enough to simply intervene; the intervention must be imbued with intelligence, as must the LA reports that trigger interventions in the first place.[16]

Fully Automated Responses

This kind of LA application delivers automatic messages in response to the findings of its analysis. Once set up, it requires little or no action on the part of instructors or advisors to initiate the intervention. The content or extent of these responses varies. They can be brief and simple nudges, such as a green/yellow/red indicator, to alert either the learner or the instructor to take a closer look.[17] At the other end of this spectrum would be the intelligent tutoring systems, which can offer help and coaching directly related to the suspected problem area.

Semi-Automated Responses

These are alerts and other indicators of significant learner patterns—often focused on downward patterns—sent to the learning "team" (some combination of instructors, students, and advisors). The patterns discovered in the LA analysis can be thought of as symptoms, and it is left to the learning team to arrive at a diagnosis and determine a course of action.

The blend of intervention types is highly dependent on institutional capacity, culture, and strategic directions. Some projects move beyond individual interventions within a single course to include ways to manage a learner's "case" across courses and semesters. This approach builds a kind of LA dossier or file that would capture past interventions, notes about the learner's circumstances, and other information that would serve to tailor subsequent interventions to that learner's circumstances.

As we have seen, the work reported by Essa and Ayad incorporates "a case-based methodology for managing interventions," analogous to the way cases are managed in the medical domain. It provides an up-to-date history of a learner's interactions with faculty, advisors, and other contacts in student services, all serving to deliver a more detailed and nuanced "portrait" of the learner.

The Point of It All

At a LAK12 general session, David Wiley of Brigham Young University reminded us of Bloom's 2 sigma problem. Benjamin Bloom discovered that with one-on-one or one-on-two tutoring, even average students could perform two standard deviations higher than average students who were taught using conventional group methods. Having ascertained this, Bloom's quest was to "find methods of group instruction as effective as one-to-one tutoring."[18]

To implement such a system of highly personalized tutoring would of course be prohibitively expensive. Wiley's point was that any system or program that we can build that moves us in that direction—that personalizes the educational experience and tailors resources to specific student needs—will enable the students to achieve more. This is, he suggests, the true potential of LA-based undertakings.

This theme was particularly evident in Tim McKay's presentation on the University of Michigan's E²Coach program (also supported by NGLC). The motivation for this LA project is to foster greater rates of degree completion in STEM disciplines. He summarized the challenge: "Students in gateway STEM courses are diverse by many measures, yet we ask them to learn using a single generic approach. They all read the same texts, hear the same lectures, do the same homework and class assignments, get the same advice, and are assessed using the same exams." The goal of the E²Coach project is "to provide individual advice and coaching to every student."[19]

One of the most interesting aspects that will be investigated by the Michigan work is whether the impact of an LA program can substantially alter past patterns. Researchers there have found—as have many others—that prior and early performance is usually a strong predictor of overall course outcome. Their investigation has "shown a disappointingly strong correlation between first exam performance and subsequent work for all students." Yet there seems to be no inherent reason for this pattern. Bloom's insight suggests that it might stem in part from the impersonal design of many gateway courses. As the Michigan team writes, "Students who change their approach to the class are likely to improve their outcomes."

The question we all share in higher education is whether a program built around LA can disrupt this pattern for some, if not many, of our students. LA and the programs we build around that technology are emergent—both will need to evolve in terms of sophistication and effectiveness. The early reports delivered at the Spring Focus Session and the LAK12 conference are encouraging. Over the coming months, ELI will continue to watch this evolution, providing events, presentations, and publications to help the community explore ways that LA-based programs can make a difference for students and faculty alike.

EDUCAUSE will maintain its focus on analytics in the upcoming months, and much of that discussion will deal with LA. Some of the important events include an Analytics Sprint (July 24–26, 2012) and a series of case studies that will be appearing in July and August of 2012. Details on these and other analytics-related activities can be found on the EDUCAUSE website.

Notes

1. See http://www.solaresearch.org/about/.
2. Veronica Diaz and Malcolm Brown, "Learning Analytics: A Report on the ELI Focus Session" (Louisville, CO: EDUCAUSE, May 2012), 3.
3. Tim McKay, Kate Miller, and Jared Tritz, "What to Do with Actionable Intelligence: E²Coach as an Intervention Engine" (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012), 2.
4. Simon Buckingham Shum and Ruth Crick, "Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics" (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012), 1; project blog at http://learningemergence.net/.
5. See http://en.wikipedia.org/wiki/Learning_power.
6. Buckingham Shum and Crick, "Learning Dispositions," 6.
7. Diaz and Brown, "Learning Analytics," 11–12.
8. Kimberly E. Arnold and Matthew D Pistilli, "Course Signals at Purdue: Using Learning Analytics to Increase Student Success," (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012), 1.
9. Mike Sharkey and Rebecca Barber, "Course Correction: Using Analytics to Predict Course Success," (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012).
10. Andrew E. Krumm, R. Joseph Waddington, Stephanie D. Teasley, and Steven Lonn, "Bridging the Gap from Knowledge to Action: Putting Analytics in the Hands of Academic Advisors," (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012), 2.
11. Alfred Essa and Hanan Ayad, "Student Success System: Risk Analytics and Data Visualization using Ensembles of Predictive Models," (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012), 1.
12. Brandon White and Jóhann Ari Lárusson, "Monitoring Student Progress through Their Written 'Point of Originality,'" (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012).
13. Erik Duval, Jose Luis Santos, Katrien Verbert, and Sten Govaerts, "Goal-Oriented Visualizations of Activity Tracking: A Case Study with Engineering Students," (paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012).
14. Essa and Ayad, "Student Success System," 3.
15. Krumm et al., "Bridging the Gap," 3.
16. See http://net.educause.edu/ELIWEB126.
17. Tim Coley, "The Case for Nudge Analytics," *EDUCAUSE Quarterly* 33, no. 4 (2010).
18. Benjamin S. Bloom, "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," *Educational Researcher* 13, no.6. (Jun.–Jul. 1984), 4–16.
19. McKay, Miller, and Tritz, "What to Do with Actionable Intelligence," 1, 3.