



ILLUSTRATION BY FRANCESCO BONGIORNI, © 2011

Penetrating the FOGG

Analytics in Learning and Education

By Phil Long and George Siemens

Attempts to imagine the future of education often emphasize new technologies—ubiquitous computing devices, flexible classroom designs, and innovative visual displays. But the most dramatic factor shaping the future of higher education is something that we can't actually touch or see: *big data and analytics*. Basing decisions on data and evidence seems stunningly obvious, and indeed, research indicates that data-driven decision-making improves organizational output and productivity.¹ For many leaders in higher education, however, experience and “gut instinct” have a stronger pull.

Meanwhile, the move toward using data and evidence to make decisions is transforming other fields. Notable is the shift from clinical practice to evidence-based medicine in health care. The former relies on individual physicians basing their treatment decisions on their personal experience with earlier patient cases.² The latter is about carefully designed data collection that builds up evidence on which clinical decisions are based. Medicine is looking even further toward computational

acades, calls have been made for reform in the efficiency and quality of higher education. Now, with the Internet, mobile technologies, and open education, these calls are gaining a new level of urgency. Compounding this technological and social change, prominent investors and businesspeople are questioning the time and monetary value of higher education.⁵ Unfortunately, the crescendo of calls for higher education reform lacks a foundation for making decisions on what to do or how to guide

vide insight into which students are at risk of dropping out or need additional support to increase their success, and confidence, in the learning process. Indeed, some in higher education have recently begun to consider how to apply analytics to better understand the learning process. EDUCAUSE and the Next Generation Learning Challenge, or NGLC (<http://nextgenlearning.org/>), are focusing the educational community on the possibilities that can be achieved by modeling learning



Higher education, a field that gathers an astonishing array of data about its “customers,” has traditionally been inefficient in its data use, often operating with substantial delays in analyzing readily evident data and feedback.

modeling by using analytics to answer the simple question “who will get sick?” and then acting on those predictions to assist individuals in making lifestyle or health changes.³ Insurance companies also are turning to predictive modeling to determine high-risk customers. Effective data analysis can produce insight into how lifestyle choices and personal health habits affect long-term risks.⁴ Business and governments too are jumping on the analytics and data-driven decision-making trends, in the form of “business intelligence.”

Higher education, a field that gathers an astonishing array of data about its “customers,” has traditionally been inefficient in its data use, often operating with substantial delays in analyzing readily evident data and feedback. Evaluating student dropouts on an annual basis leaves gaping holes of delayed action and opportunities for intervention. Organizational processes—such as planning and resource allocation—often fail to utilize large amounts of data on effective learning practices, student profiles, and needed interventions.

Something must change. For de-

change. It is here—as a framework for making learning-based reform decisions—that analytics will have the largest impact on higher education.

Data Explosion

A byproduct of the Internet, computers, mobile devices, and enterprise learning management systems (LMSs) is the transition from ephemeral to captured, explicit data. Listening to a classroom lecture or reading a book leaves limited trails. A hallway conversation essentially vaporizes as soon as it is concluded. However, every click, every Tweet or Facebook status update, every social interaction, and every page read online can leave a digital footprint. Additionally, online learning, digital student records, student cards, sensors, and mobile devices now capture rich data trails and activity streams.

These learner-produced data trails provide valuable insight into what is actually happening in the learning process and suggest ways in which educators can make improvements. Analysis of learner data may also pro-

vide insight into which students are at risk of dropping out or need additional support to increase their success, and confidence, in the learning process.

The idea is simple yet potentially transformative: analytics provides a new model for college and university leaders to improve teaching, learning, organizational efficiency, and decision making and, as a consequence, serve as a foundation for systemic change. But using analytics requires that we think carefully about what we need to know and what data is most likely to tell us what we need to know. Continued growth in the amount of data creates an environment in which new or novel approaches are required to understand the patterns of value that exist within the data. P.W. Anderson stated that “more is different,” emphasizing that new models of and approaches to data interaction are desperately needed when we are confronted with abundance. Or, as stated by David Gelernter: “If you have three pet dogs, give them names. If you have 10,000 head of cattle, don’t bother.”⁶ Quantity changes the methods and approaches that we use to interact with and make sense of data.

Google’s Marissa Mayer⁷ suggests that data is today defined by three elements:

1. *Speed*—The increasing availability of data in real time, making it possible to process and act on it instantaneously
2. *Scale*—Increase in computing power: Moore’s law (stating that the number of transistors on a circuit board will double roughly every two years) continues to hold true.
3. *Sensors*—New types of data: “Social data is set to be surpassed in the data economy, though, by data published by physical, real-world objects like sensors, smart grids and connected devices”—that is, the “Internet of Things.”⁸

Taken together, these three elements create a situation in which existing data-management and decision-making approaches simply are not feasible. Understanding how activities such as research, teaching, and support services contribute to learners’ achievement is not possible in the currently largely linear data-collection and data-analysis model. Information abundance, and the attendant institutional complexity involved in defining and enacting strategy, suggest rethinking the role that analytics can play in making sense of data.

Big Data

Big data is a term used to describe the new context of abundance. The McKinsey Global Institute defines big data as “datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze.”⁹ In response to the limitations of existing data-management techniques, a new breed of technologies (e.g., Hadoop), databases, and techniques (e.g., data-mining or knowledge discovery in databases) has been developed. As a consequence, theorists have posited that something fundamental has changed with the data itself, creating a

world in which almost all data interactions, including scientific research, are affected:

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.¹⁰

The key emphasis in big data is that the data itself is a point of or a path to value generation in organizations. Data is not simply the byproduct of interactions and activities within an organization. Data is a critical value layer for governments, corporations, and higher education institutions.

Learning Analytics

In colleges and universities, the data focus is increasingly expressed using the term *learning analytics*. Though still a young concept in education, learning analytics already suffers from term sprawl. The ubiquity of the term *analytics* partly

contributes to the breadth of meanings attached to it. For our purposes here, a reasonable definition of *learning analytics* will help to guide discussion and frame activities.

According to the 1st International Conference on Learning Analytics and Knowledge, “learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.”¹¹ Academic analytics, in contrast, is the application of business intelligence in education and emphasizes analytics at institutional, regional, and international levels. John P. Campbell, Peter B. DeBlois, and Diana G. Oblinger stated: “Analytics marries large data sets, statistical techniques, and predictive modeling. It could be thought of as the practice of mining institutional data to produce ‘actionable intelligence.’”¹²

Learning analytics is more specific than academic analytics: the focus of the former is exclusively on the learning process, as detailed in Table 1. Academic analytics reflects the role of data analysis at an institutional level, whereas learning analytics centers on the learning process (which includes analyzing the relationship between

TABLE 1: LEARNING AND ACADEMIC ANALYTICS

TYPE OF ANALYTICS	LEVEL OR OBJECT OF ANALYSIS	WHO BENEFITS?
Learning Analytics	Course-level: social networks, conceptual development, discourse analysis, “intelligent curriculum”	Learners, faculty
	Departmental: predictive modeling, patterns of success/failure	Learners, faculty
Academic Analytics	Institutional: learner profiles, performance of academics, knowledge flow	Administrators, funders, marketing
	Regional (state/provincial): comparisons between systems	Funders, administrators
	National and International	National governments, education authorities



Academic analytics reflects the role of data analysis at an institutional level, whereas learning analytics centers on the learning process (which includes analyzing the relationship between learner, content, institution, and educator).

learner, content, institution, and educator).

The distinction of academic analytics as similar to business intelligence raises the need for a model or stage of learning analytics development. We propose the following cycle to reflect analytics in learning:

1. *Course-level*: learning trails, social network analysis, discourse analysis
2. *Educational data-mining*: predictive modeling, clustering, pattern mining
3. *Intelligent curriculum*: the development of semantically defined curricular resources
4. *Adaptive content*: adaptive sequence of content based on learner behavior, recommender systems
5. *Adaptive learning*: the adaptive learning process (social interactions, learning activity, learner support, not only content)

The Value of Analytics for Higher Education

Analytics spans the full scope and range of activity in higher education, affecting administration, research, teaching and learning, and support resources. The college/university thus must become a more intentional, intelligent organization, with data, evidence, and analytics playing the central role in this transition.

How do big data and analytics generate value for higher education?

1. They can improve administrative decision-making and organizational resource allocation.
2. They can identify at-risk learners and provide intervention to assist

learners in achieving success. By analyzing discussion messages posted, assignments completed, and messages read in LMSs such as Moodle and Desire2Learn, educators can identify students who are at risk of dropping out.¹³

3. They can create, through transparent data and analysis, a shared understanding of the institution's successes and challenges.
4. They can innovate and transform the college/university system, as well as academic models and pedagogical approaches.
5. They can assist in making sense of complex topics through the combination of social networks and technical and information networks: that is, algorithms can recognize and provide insight into data and at-risk challenges.
6. They can help leaders transition to holistic decision-making through analyses of what-if scenarios and experimentation to explore how various elements within a complex discipline (e.g., retaining students, reducing costs) connect and to explore the impact of changing core elements.
7. They can increase organizational productivity and effectiveness by providing up-to-date information and allowing rapid response to challenges.
8. They can help institutional leaders determine the hard (e.g., patents, research) and soft (e.g., reputation, profile, quality of teaching) value generated by faculty activity.¹⁴
9. They can provide learners with insight into their own learning habits and can give recommendations

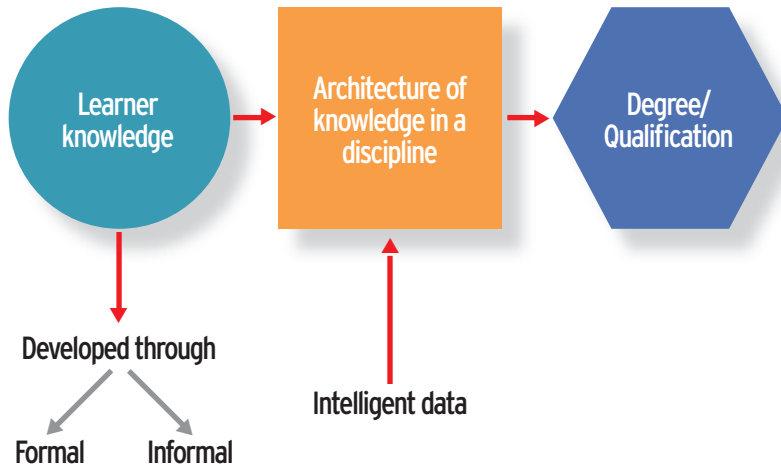
for improvement. Learning-facing analytics, such as the University of Maryland, Baltimore County (UMBC) Check My Activity tool, allows learners to “compare their own activity . . . against an anonymous summary of their course peers.”¹⁵

Moving Beyond the LMS

Analytics from LMSs—or VLEs (Virtual Learning Environments), as they are known in Europe)—offers one source of data for predicting the success of learners. Morris, Finnegan, and Wu compared basic activities related to LMS participation (e.g., content pages viewed, number of posts) and duration of participation (e.g., hours spent viewing discussion pages and content) in LMSs and found significant differences between “withdrawers” and “successful completers,” concluding that “time spent on task and frequency of participation are important for successful online learning.”¹⁶ Leah P. Macfadyen and Shane Dawson advocate for early-warning reporting tools that “can flag at-risk students and allow instructors to develop early intervention strategies.”¹⁷

LMSs have been adopted as learning analytics tools because the data captured is structured and reflects the learners' interaction within a system. But distributed networks and physical world interactions present additional challenges for analytics. For example, most LMS analytics models do not capture activity by online learners outside of an LMS (i.e., in Facebook, Twitter, or blogs). Similarly, most analytics models do not capture or utilize physical-world data, such as library use, access to learning support, or academic advising.

FIGURE 1: ASSESSMENT THROUGH ANALYTICS



Mobile devices such as smartphones and tablets/iPads offer the prospect of bridging the divide between the physical and digital worlds by capturing location and activity. Similarly, clickers in classrooms can be integrated with data from learners' activity in online environments, providing additional insight into factors that contribute to learners' success.

Massive Open Online Courses (MOOCs), which occur in decentralized, distributed teaching and learning networks, offer another challenge. Online social media monitoring tools (e.g., Radian6) and reputation or influence monitoring tools (e.g., Klout) may provide educators with a model for analytics in such networks, in which activity is distributed across multiple sites and multiple identities.

Intelligent Curriculum

It is not sufficient to treat big data and analytics as useful only for evaluating what learners have done and for pre-

dicting what they'll do in the future. Analytics in education must be transformative, altering existing teaching, learning, and assessment processes, academic work, and administration.

When analytics is applied to curricular resources, the traditional view of courses is disrupted. The knowledge, attitudes, and skills required in any domain can be rendered as a network of relations. The semantic web and linked data are partial instantiations of this concept. Knowledge domains can be mapped, and learner activity can be evaluated in relation to those maps. Instead of being an "end of course" activity, assessment is performed in real time as learners demonstrate mastery of important concepts or ideas (see Figure 1). Learning content is not provided in a packaged textbook but is rendered or computed "on the fly," providing each learner with resources relevant to his or her profile, learning goals, and the knowledge domain the learner is attempting to master. This is the essence

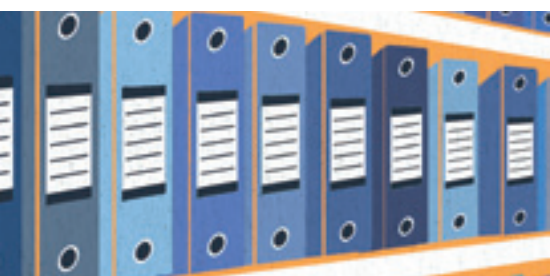
of the success that has accompanied the development of the Khan Academy learning modules, even with their simplistic, mastery-based approach.¹⁸

Concluding Thoughts

Learning analytics is still in the early stages of implementation and experimentation. Numerous questions exist around how analytics relates to existing organizational systems. Campbell, DeBlois, and Oblinger detailed the various concerns that the use of analytics generates in higher education, including privacy, profiling, information sharing, and data stewardship.¹⁹ How can the potential value of the data be leveraged without succumbing to the dangers associated with tracking students' learning options based on deterministic modeling? Additionally, how transparent are the algorithms and weighting of analytics? How "real time" should analytics be in classroom settings? Finally, since we risk a return to behaviorism as a learning theory if we confine analytics to behavioral data, how can we account for more than behavioral data?

Undoubtedly, analytics and big data have a significant role to play in the future of higher education. The growing role of analysis techniques and technologies in government and business sectors affirms this trend. In education the value of analytics and big data can be found in (1) their role in guiding reform activities in higher education, and (2) how they can assist educators in improving teaching and learning.

Yet there are reasons to be cautious as the development of analytical tools for modeling learners' interactions



It is not sufficient to treat big data and analytics as useful only for evaluating what learners have done and for predicting what they'll do in the future. Analytics in education must be transformative.

gains attention. Like other behavior patterns, models that are deterministic assume that future conditions can be completely determined by knowing both the past and the present conditions of the subject involved. This can be a convenient simplification from the more challenging requirements of alternative approaches. Stochastic models, on the other hand, are probabilistic: even with full knowledge of the present state of things, we cannot be sure of the future. We must guard against drawing conclusions about learning processes based on questionable assumptions that misapply simple



models to a complex challenge. Learning is messy, and using analytics to describe learning won't be easy.

Learning analytics is essential for penetrating the fog that has settled over much of higher education. Educators, students, and administrators need a foundation on which to enact change. For educators, the availability of real-time insight into the performance of learners—including students who are at-risk—can be a significant help in the planning of teaching activities. For students, receiving information about their performance in relation to

their peers or about their progress in relation to their personal goals can be motivating and encouraging. Finally, administrators and decision-makers are today confronted with tremendous uncertainty in the face of budget cuts and global competition in higher education. Learning analytics can penetrate the fog of uncertainty around how to allocate resources, develop competitive advantages, and most important, improve the quality and value of the learning experience. ■

Notes

1. Erik Brynjolfsson, Lorin M. Hitt, and Heekyung Hellen Kim, "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" *Social Science Research Network*, Working Paper Series, April 22, 2011, <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486#>.
2. See "Evidence-based Medicine: What Does It Really Mean?" *Progress in Reproductive Health Research*, 1999, <http://www.reproline.jhu.edu/english/6read/6issues/6progress/prog54_b.htm>.
3. See Heritage Provider Network Health Prize Competition: <<http://www.heritagehealthprize.com/c/hhp>>.
4. Leslie Scism and Mark Maremont, "Insurers Test Data Profiles to Identify Risky Clients," *Wall Street Journal*, November 19, 2010, <<http://online.wsj.com/article/SB10001424052748704648604575620750998072986.html>>.
5. Sarah Lacy, "Peter Thiel: We're in a Bubble and It's Not the Internet, It's Higher Education," *TechCrunch*, April 10, 2011, <<http://techcrunch.com/2011/04/10/peter-thiel-were-in-a-bubble-and-its-not-the-internet-its-higher-education/>>; William H. Gross, "School Daze, School Daze, Good Old Golden Rule Days," PIMCO, <<http://www.pimco.com/EN/Insights/Pages/School-Daze-School-Daze-Good-Old-Golden-Rule-Days.aspx>>.
6. P. W. Anderson, "More Is Different," *Science*, vol. 177, no. 4047 (August 4, 1972), pp. 393–396; David Gelernter, "The Second Coming: A Manifesto," *Edge*, December 31, 1999, <<http://edge.org/conversation/the-second-coming-a-manifesto>>.
7. "Innovation at Google: The Physics of Data," PARC Forum, <<http://www.slideshare.net/PARCInc/innovation-at-google-the-physics-of-data>>.
8. Marshall Kirkpatrick, "China Moves to Dominate the Next Stage of the Web," *ReadWriteWeb*, August 12, 2010, <http://www.readriteweb.com/archives/china_moves_to_dominat_the_next_stage_of_the_web_internet_of_things.php>.
9. James Manyika, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," Executive Summary, McKinsey Global Institute, May 2011, <http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_exec_summary.pdf>.

10. Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, June 23, 2008, <http://www.wired.com/science/discoveries/magazine/16-07/pb_theory>.
11. 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, February 27–March 1, 2011, <<https://tekri.athabasca.ca/analytics/>>.
12. John P. Campbell, Peter B. DeBlois, and Diana G. Oblinger, "Academic Analytics: A New Tool for a New Era," *EDUCAUSE Review*, vol. 42, no. 4 (July/August 2007), pp. 40–57, <<http://www.educause.edu/library/erm0742>>.
13. Leah P. Macfadyen and Shane Dawson, "Mining LMS Data to Develop an 'Early Warning System' for Educators: A Proof of Concept," *Computers & Education*, vol. 54, no. 2 (2010), pp. 588–599.
14. The approach of determining value remains controversial, since many aspects of the educational system do not map to economic value. See Simon Head, "The Grim Threat to British Universities," *New York Review of Books*, December 16, 2010, <<http://www.nybooks.com/articles/archives/2011/jan/13/grim-threat-british-universities/?page=1>>.
15. John Fritz, guest speaker, "Introduction to Learning and Knowledge Analytics: An Open Online Course," week 1, January 11, 2011, <<http://www.learninganalytics.net/syllabus.html>>.
16. Libby V. Morris, Catherine Finnegan, and Sz-Shyan Wu, "Tracking Student Behavior, Persistence, and Achievement in Online Courses," *The Internet and Higher Education*, vol. 8, no. 3 (2005), pp. 221–231.
17. Macfadyen and Dawson, "Mining LMS Data," p. 589.
18. See Clive Thompson, "How Khan Academy Is Changing the Rules of Education," *Wired*, July 15, 2011, <http://www.wired.com/magazine/2011/07/ff_khan/all/1>.
19. Campbell, DeBlois, and Oblinger, "Academic Analytics."

© 2011 Phil Long and George Siemens. The text of this article is licensed under the Creative Commons Attribution-NonCommercial 3.0 License (<http://creativecommons.org/licenses/by-nc/3.0/>).



Phil Long (longpd@uq.edu.au) is a Professor in the Schools of ITEE and Psychology and is Director of the Centre for Educational Innovation & Technology at the University of Queensland.



George Siemens (gsiemens@gmail.com) is with the Technology Enhanced Knowledge Research Institute at Athabasca University.