An Overview of Learning Analytics

UCLA Office of Information Technology
Jim Williamson
September 2014
# Table of Contents

**Executive Summary** ................................................................. 3  
**Introduction** ............................................................................ 4  
**Defining “Learning Analytics”** .................................................. 4  
  *Table: Levels of Useful Data for Understanding Student Performance* ........................................ 5  
**Goals of Learning Analytics** .................................................... 6  
**Types of Learning Analytics** .................................................... 7  
  A. Evidence of Student Progress .................................................. 7  
  *Future Trends* ......................................................................... 7  
  B. Evidence of Student Behavior .................................................. 7  
  *Example: Purdue Signals* ......................................................... 7  
  *Example: Learning Catalytics* ................................................... 7  
  *Example: Open Academic Analytics Initiative* ......................... 7  
  *Future Trends* ......................................................................... 7  
  C. Evidence of Student Learning .................................................. 8  
  *Example: CRESST Work* ......................................................... 8  
  *Example: Khan Academy* ....................................................... 8  
  *Example: Saint-Petersburg State University of Information Technologies, Mechanics, and Optics* .......... 9  
  *Future Trends* ......................................................................... 9  
**The Range of Learning Analytics** ............................................. 9  
**Learning Analytics Data Frameworks and Repositories** ............ 9  
  *Example: Predictive Analytics Reporting (PAR) Framework* ...................................................... 10  
  *Example: Advance Distributed Learning Experience API/Tin Can API* ..................................... 10  
  *Example: IMS Global Caliper Framework* .............................................................................. 10  
  *Example: Pittsburgh Science of Learning Center DataShop* ..................................................... 10  
  *Future Trends* ......................................................................... 11  
**Privacy, Ownership, and Policy** .............................................. 11  
**Projections and Opportunities** .................................................. 13
Executive Summary

“Learning analytics” refers to a wide set of data gathering, storing, and reporting techniques for administrative, programmatic, and pedagogical purposes. The information from this data ranges from measuring retention and student progress, to course tool usage, to very granular, personalized, specific student data.

The sources for this data are equally wide, ranging from student information systems, learning management systems, or task-specific learning tools. The reporting from this data is descriptive and predictive, and leverages a wide variety of tools, from enterprise business tools to custom algorithms.

The emergence of learning analytics parallels the nation-wide interest in evidenced-based learning and affordances from digital delivery of learning content. It remains to be seen if and how pedagogical value, technological complexities, and cultural barriers will impact the adoption of learning analytics.
An Overview of Learning Analytics

Introduction

Data in an educational setting is certainly not new. For decades, researchers, administrators, and other interested stakeholders at the institutional level have used educational data to evaluate and compare institutions and programs, report graduation and retention rates, and make enrollment and resource projections. At the learner level, data from intelligent agents, learning objects, games, and simulations has enabled scientists to inform their pedagogical and cognitive research.

Although educational data is not new, the term “learning analytics” has emerged in the past few years as an important focal point for higher education and is often driven by language such as, “in the service of improving learning and education”¹ or evaluating “the effectiveness of online instruction in delivering a UC-quality undergraduate education.”² Now, administrators are using business intelligence tools to project student recruitment and retention trends. Faculty are using LMS data to gain new insights into their classes and students. Vendors are entering this market in the hopes of providing services unavailable to institutions. Politicians, business leaders, and accreditors seek evidence of student learning to ensure quality and justify the public investment. Indicators of this new focus on learning analytics include: the publication of first issue of the Journal of Educational Data Mining (JEDM) (2009); the Horizon Report’s identification of learning analytics as a technology to watch for (2011); and the first conference of the Society of Learning Analytics Research (SoLR) (2011). Products such as Knewton, Signals, and Cognitive Tutor from vendors such as Cengage and Pearson, and universities like Purdue and Carnegie Mellon are now available in the marketplace.

Even more recently, in just the first six months of this year, we have seen the emergence of the IMS Global Caliper Project, the EDUCAUSE ICAR Integrated Planning and Advising Services project, and the call from the Institute for Higher Education Policy to answer the “million dollar question”:

> The million-dollar question facing all higher education constituencies involves the concept of “value.” Which colleges produce value or return on investment for students, as measured by inputs, outputs, and cost?³

The current thriving focus on learning analytics can be attributed to the availability, scale, and granularity of educational data from multiple new sources being processed through business intelligence tool and predictive statistical methods aiming toward a variety of new goals and insights. Contemporary learning analytics encompasses new technologies, applications and reporting methods, descriptive and predictive methods, data mining, and accountability pressures from inside and outside the academy.

Defining “Learning Analytics”

Since this document is intended as a broad overview, we use the broad term “learning analytics” to encompass educational data for a variety of goals and from a variety of sources. More specific definitions are emerging, although the terminology has not yet entirely settled. For example, Siemens and Long divide educational data into two categories and five levels of scope: Learning Analytics (Course and Departmental) and Academic Analytics (Institutional, Regional and National/International):

---

² UCOP. “The UC Online Instruction Pilot Project Request for Letters of Intent from UC Academic Senate faculty” (2010)
“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.” Academic analytics, in contrast, is the application of business intelligence in education and emphasizes analytics at institutional, regional, and international levels. Greg Chung, a researcher at UCLA's CRESST, divides learning analytics between a coarse-grained learning analytics approach and a fine-grained “learning process” approach.

To-date, learning analytics has focused primarily on the analysis of the behavior and interaction of students inside the K-12 or training environments where computer games and simulations directly impact learning. These types of content delivery differ from the modes of learning found in the majority of traditional higher education, but research is emerging around higher educational course work where learning outcomes and directed pathways play a role, such as in STEM and medicine.

Table: Levels of Useful Data for Understanding Student Performance

<table>
<thead>
<tr>
<th>Data Level</th>
<th>Data Source</th>
<th>Primary User</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>System-level</td>
<td>SIS; LMS</td>
<td>Institution</td>
<td>Course grades, high school records, student demographics</td>
</tr>
<tr>
<td>Individual-level</td>
<td>Quiz; assignments; eBooks; media player</td>
<td>Student, faculty</td>
<td>Achievement test scores; overall quiz result; response to individual items in a test</td>
</tr>
<tr>
<td>Transaction-level</td>
<td>Game or simulation interaction</td>
<td>Faculty, researcher, developer</td>
<td>Captures a student’s moment-by-moment performance on a task, such as mouse coordinates or click rate; use of hints or help system</td>
</tr>
</tbody>
</table>

Descriptive analytics are based on static data derived from a number of sources, such as course evaluations, student exit surveys, student information systems, LMS activity, and ePortfolio interactions. The subsequent reporting may span both quantitative and qualitative data, and the nature of this data is post hoc and often summative. Examples include the quarterly UCLA CCLE report which provides an overview of the LMS system and tool use, or the qualitative review of the Honors Collegium 70A from the UCLA Center for Educational Assessment which uses open-ended feedback, course exit questionnaires, and interviews with former students.

Predictive analytics may extend data from the same sources, but focuses on trying to measure actual learning. The data may come from intelligent agents, task-specific games, log files, simulations designed to capture the learning process, and/or direct observation. Measuring learning is notoriously difficult, and researchers continue to refine data and methods to achieve insights into the learning process.

Whereas descriptive and predictive analytics involve asking questions and receiving answers, data mining is an exploratory process of discovery. Researchers develop and apply algorithms and models to data sets about the learning environment to reveal previously unseen patterns of information. One of the attractions of the large-scale MOOC courses is the hope of unearthing new and surprising information about how people learn.

---

Moving beyond the educational setting, there are other aspects to this data that are emerging as important drivers or potentially interesting research. First, Tom Gilbert notes that contracts with video providers required specific data about student consumption: "One of the biggest drivers ... was how long are people watching these videos? We have royalty payments and contractual rights that we are obligated to report back out to our authors and content providers. That was non-negotiable for us: we had to track it to the minute and to the second ...".

Second, researchers like Simon Buckingham Shum are exploring Social Learning Analytics: "As ubiquitous access to social networks become a critical part of learners’ online identity, and an expected part of learning platforms, social learning analytics should provide tools to provoke learning-centric reflection on how interpersonal relationships and interactions reflect learning, or can be more effectively used to advance learning."

Capturing atomic values from demographics, college applications, social media, and computer-based, digital course content delivery will allow researchers to reconsider how to evaluate effective learning across a student’s entire digital environment. In what might be termed “holistically quantitative,” the data focus slides across a continuum from abstract, summative, theme-based course assessment to detailed granular analysis of an individual student’s knowledge, skills, and attributes (KSAs). This shift is well underway in K-12 (due to reliance on Common Core standards) and training realms, and as will be discussed, depending on the discipline, it should not surprise us to find these measurement techniques used and expected in higher education.

**Goals of Learning Analytics**

Learning analytic goals range from high-level university administrative functions such as cohort retention, time-to-degree and resource allocation, to observing and recording the amount of distraction for 4th grade male students working on a classroom math problem, to measuring a university student’s heart rate when taking an exam.

Because of the diversity of stakeholders, goals, data types and sources, it is important to ensure that the broad term “learning analytics” is properly contextualized and understood at each use. For example, the Open Academic Analytics Initiative (OAAI) project web page states, “By ‘learning analytics’ we mean ‘the use of analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals.’ But understanding this definition requires understanding what OAAI means by “specific learning goals.” In their context, learning goals equate to “increase college student retention.”

The majority of what is labeled “learning analytics” is not actually about student learning--which is admittedly difficult to measure--but is often something used as a proxy for learning, such as student activity. Student retention, time-to-degree, and resource allocation are all worthy pursuits, but activity should not be confused with learning. We stress this point so that readers do not mistakenly pursue a path or follow claims about learning analytics that will not prove beneficial to their goals.

Thus, there is a need to embrace the spectrum of learning analytics. While not always focused on individual student learning, studies have illustrated that quality instruction emerges from institutions who use data to focus on student success: “decision-making processes relating to organization of institutional resources –

---

9 Baron, Joshua D. 'Scaling Learning Analytics across Institutions of Higher Education': EDUCAUSE, SEI Case Study. (2013)
human and material – and planning for more effective use of existing resources are a critical feature of excellent institutions.”

Types of Learning Analytics

Educational data falls into three broad and at times, overlapping, categories of discovery: student progress, student behavior, and student learning. Introducing external linked data, such as demographic or social media sources, can augment each of these approaches.

A. Evidence of Student Progress

- High-level data
- Generally found at the institution level
- The most traditional and mature data gathering and reporting in higher education
- Sample sources: SIS, Registrar, and other institutional data stores
- Data includes: SAT scores, high school GPA, final course grades, degree progress markers

Future Trends

Linked data will enable connecting high school scores, demographic information from college entrance applications, financial aid applications, and targeting social media sources.

B. Evidence of Student Behavior

- Mid-level data
- Generally focused on student achievement, such as course activity and grades
- Sample sources: SIS, LMS, data collection from tools such as clickers
- Data includes: web server log information, LMS or other course tool activity

Example: Purdue Signals

http://www.itap.purdue.edu/studio/signals/

“To identify students at risk academically, Signals combines predictive modeling with data-mining from Blackboard Vista. Each student is assigned a "risk group" determined by a predictive student success algorithm. One of three stoplight ratings, which correspond to the risk group, can be released on students’ Blackboard homepage.”

Example: Learning Catalytics

https://learningcatalytics.com

Created by Harvard faculty Eric Mazur and Gary King and later acquired by Pearson. Uses web-enabled devices such as laptop, smartphone, or tablet to engage students by creating open-ended questions that ask for numerical, algebraic, textual, or graphical responses or multiple-choice.

Example: Open Academic Analytics Initiative

https://confluence.sakaiproject.org/pages/viewpage.action?pageId=75671025

OAAI has structured their analytics system on “student performance” defined as course completion, content mastery and semester-to-semester persistence. Their coarse-grained data sources included:

- Student Demographics Data (Personal information)
- Course Data

• LMS Gradebook Data
• LMS Usage/Events Data

Future Trends
Because this sector is the most interesting to university administrators, there is a great deal of vendor activity in this area and may become the focal point for most funding. However, the experience of LSU and their Early Academic Referral System (EARS) is worth noting:

We spent months in development, but, due to some political kickback from faculty, we were forced to limit the scope of the system to individual courses (to prevent any possibility of longterm, longitudinal data gathering). We were also instructed to make the system “opt-in” as opposed to “opt-out”. In the end, after all the work was done and the system was vetted and approved by the administration, we rolled it out with some amount of fanfare. After several semesters, we had exactly one faculty member using it.

C. Evidence of Student Learning

• Fine-grain data
• Application-specific data: recording learning events in real time
• Combines psychometric techniques, LMS data, and adaptive systems, such as intelligent agents
• Focused on personalized student learning
• Sample source: Games, simulations, intelligent agents

Example: CRESST Work
The UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST) divides the learning analytics spectrum between outcome measures and process measures, and attempts to gather information through student activity in games and simulations to address three aspects of cognition. “Underlying all three questions is the collection of fine-grained behavioral data and its use to gain insight about students’ learning processes.”

a) To what extent does students’ online behavior relate to their cognitive processing?
b) To what extent can students’ online behavior be used to model their problem solving process?
c) To what extent can students’ online behavior be used diagnostically to reveal misunderstandings and misconceptions?

CRESST research “found significant correlations between students’ online behavior and their cognitive processing, particularly with processes that required reasoning (e.g., drawing correct inferences from a series of test results).”

Example: Khan Academy
A predictive learning analytics model from the Khan Academy provides this scenario: Students work in a data-rich environment that includes over 100 features and student characteristics, including demographic information, past task attempts and scores (or, KSA – a student’s knowledge, skills, and attributes).

Based on Khan knowledge of the student, the learning system, and the difficulty of the student’s current learning activity, Khan attempts to predict the student’s response on the activity. Not only does the Khan model look at whether the student answered correctly or not, but it also factors in how quickly the student achieved the right or wrong answer. This not only provides information on the student’s ability to perform the activity, but also adds an efficiency marker, which helps the system guide the student toward appropriate

---

11 Open Academic Analytics Initiative: https://confluence.sakaiproject.org/pages/viewpage.action?pageId=75671025
new activities. Students are therefore not taxing their brain being tested on materials for which they have already demonstrated mastery.

Using a model that correctly predicts a student’s response to a new task based on past and current student characteristics and environmental factors, the Khan Academy suggests that the model may stand as a proxy for actual student comprehension.14

**Example: Saint-Petersburg State University of Information Technologies, Mechanics, and Optics**
Researchers from Russia devised a way to connect students, an electrocardiograph, and the LMS to create a mathematical model based on analysis of heart rate variability. “During the student’s work in the learning management system not only his answers are analyzed, but also the intension of the physiological system, determined by the method of heart rate variability analysis. Special software analyzes heart rate variability and transmits the results of analysis in the learning management system, where they are used in the algorithms of adaptive e-learning.”15

**Future Trends**
Because this is the most specific and challenging sector of learning analytics, researchers will continue to attempt to understand the cognitive processes of students.

**The Range of Learning Analytics**

- Academic planning: Shifts in demographic/discipline popularity
- Enrollment patterns: Campus/school/department/course sequence/course
- Accredited disciplines: Engineering
- Predictor of student success in a course: Purdue Signals project
- Final course grades: Recent OID CEA report
- Overall LMS usage data: CCLE quarterly report
- Specific educational tool data: CCLE, Ohmage, clickers
- Personalized learning analysis: at-risk students or peer comparisons
- Bridging student goals; departmental goals; career goals
- Measuring confidence of learning; Khan Academy, heart rate work
- Academic integrity: Biometrics or keystroke rate for academic integrity

**Learning Analytics Data Frameworks and Repositories**

Given the examples above, we see that learning data and analytics: exists as a spectrum of data types; is mined for a variety of goals by a variety of stakeholders; and currently resides in separate data silos. The next steps towards greater understanding about students and learning brings these disparate sources together into a learning data repository.

The massive data of MOOCs has brought new attention and talent into the educational sector by researchers interested in mining the data for new insights. For example, computer scientists working on artificial intelligence see the potential in collaborating with educators to create rich experimental environments to use data to explore human cognition. In the near future, we will see increased focus on developing learning contexts that provide increasingly granular data points.

There are two major elements to building a learning analytic repository: first, a framework to process the data, and second, a backend to store that data.

---

Example: Predictive Analytics Reporting (PAR) Framework
http://wcet.wiche.edu/par

“The Predictive Analytics Reporting (PAR) Framework is a multi-institutional data mining collaborative that brings together 2 year, 4 year, public, proprietary, traditional, and progressive institutions to collaborate on identifying points of student loss and to find effective practices that improve student retention in U.S. higher education. With sixteen member institutions, over 1,700,000 anonymized student records and 8,100,000 institutionally de-identified course level records, the PAR Framework offers educational stakeholders a unique multi-institutional lens for examining dimensions of student success from both unified and contextual perspectives.”

Example: Advance Distributed Learning Experience API/Tin Can API
http://www.adlnet.gov/tla/experience-api/background-and-history/
http://tincanapi.com/overview/

The Experience API (aka, xAPI, or Rustici’s commercial implementation, the Tin Can API) is a specification that describes an interface and the storage / retrieval rules that developers can implement to create a learning experience tracking service. The service works by allowing statements of experience (typically learning experiences, but could be any experience) to be delivered to and stored securely in a Learning Record Store (LRS). The format of these statements is based on Activity Streams(<Actor, Verb, Object> or “I did this.”). In this format, the Actor is the agent the statement is about, like a learner, mentor, teacher, or group. The verb describes the action of the statement, such as read, passed, or taught. And the object is what the Actor interacted with, like a book, a test, or a class.16

Example: IMS Global Caliper Framework
http://www.imsglobal.org/caliper/

Caliper seeks to standardize learning activities that measure user clicks, bookmarks, and other mid-level data from video, discussion forums, web content, or quizzes. This system captures learning activities to provide a general overview of what a student has done. As of June 2014, it is still in development.

Michael Feldstein provides sample questions for this kind of data repository:

• How many responses did Ann’s blog posts for Intro to Linguistics generate?
• Which students in the class write the posts that generate the most responses?
• What percentage of responses occurred as comments on the same page as the blog posts, and what percentage was blog posts on commenter’s own site?
• Which classes have the most activity of students responding to other students?
• What is the correlation between levels of student responses to each other and outcomes?
• What is the moment in the course when students started responding more to each other and less to the teacher?17

Example: Pittsburgh Science of Learning Center DataShop
http://learnlab.org/technologies/datashop/

At a finer-grained level, Carnegie Mellon’s Open Learning Initiative has partnered to develop the Pittsburgh Science of Learning Center DataShop. The website describes the venture as:

---

A data repository and analysis service for the learning science community. The PSLC DataShop provides two main services to the learning science community:

- a central repository to secure and store research data; and
- a set of analysis and reporting tools

Data collection and retrieval is a three-step process:

1. raw data logs are collected either in real-time or after the study;
2. data are extracted, anonymized, and stored in an analysis database; and
3. reports and queries are rendered via the web upon request.

Researchers can rapidly access standard reports such as learning curves, as well as browse data using the interactive web application. To support other analyses, the DataShop can export data to a variety of formats that can then be used in statistical software and other analysis packages.

At UCLA, CRESST researchers are proposing the development of WorldX, an open source repository that would store a range of data from administrative high and mid-level data to the fine-grained atomic data that fuels learning research.

**Future Trends**

The interest in learning analytics is very dynamic. These are the earliest days of definition, risk and opportunity, and it remains to be seen how the field will shape itself and find ways to consolidate efforts. For example, the xAPI and Caliper projects appear to be working in nearly identical ways based on nearly identical protocols. When Caliper was introduced, George Siemens noted, “I’m curious, if IMS wants to provide guidance and leadership, why they opted to go forward without much regard for existing LA activities or universities that have already made big strides in LA adoption.”

Much of the work in learning analytics frameworks and data storage remains in conceptual, proof-of-concept and pre-beta form, and with the academic cultural challenges, solid projections about the future are speculative at best.

**Privacy, Ownership, and Policy**

Learning analytics brings considerable challenges for data control, privacy, and data ownership. As institutions work to define privacy standards and policies for their students and faculty, understanding the implicit and explicit motivations of interested parties will be an important part of the discussion.

For example, the new Unizen project is comprised of three elements: educational content, a software platform, and analytics. The project proponents suggest a threat looms for institutions who do not align with their vision of higher educational technology, claiming that “isolated, single campus-based approaches” would shift control and economic power to entities outside of academia that develop and own technologies and services at scale. These owners would be able to assert the terms for use of content and data on their platforms and would almost certainly add new costs to universities. In that shift of decision rights, long held faculty and student rights regarding control of intellectual property and privacy might no longer be decided in the Academy.

---

19 http://unizin.org/2014/06/why-unizin/
Reviewing these Unizen platform components, we note that both scalable, sharable content already exist in the MERLOT and Connexions projects, and scalable, interconnected LMS networks already exist in the form of Moodle Community Hub. The one component that has not yet matured is the area that is also attracting the recent elevated level of interest: analytics.

(It might be noted that the software platform component of Unizen—Instructure’s Canvas—is provided as Software-As-A-Service (SAAS) which means that institutions participating in the project will be not have direct control over their cloud-hosted data and courses, so rights regarding control of intellectual property and privacy might no longer be decided in the Academy.)

Even with policies intended to safeguard and provide guidance, external service providers will increasingly have direct and indirect access to previously isolated personal data of faculty and students. The IMS Global LTI and Caliper projects emerge from public-private partnerships driven by Oracle, Pearson, Desire2Learn and other publishers and third-party tool providers. Both the EDUCAUSE ECAR IPAS project and the Institute for Higher Education Policy are funded in part by the Gates Foundation, whose position, influence and funding regularly calls for upending the staid practices of academia. Further, institutions will soon be able to leverage data mining tools from Microsoft’s Azure, cloud system:

Azure Machine Learning is designed to make this kind of analytic and processing capability available to a much wider range of organizations, with graphical workflow builders and templates to make it easier to set up data processing workloads and algorithms developed for—and proven by—services such as Bing and Xbox Live.

The features and availability of these kinds of services will overwhelm the capacity and capabilities of university IT departments to offer equivalent solutions or access, so economic and practical necessity dictate that higher education will increasingly pivot from being a service provider to being a service user. Locally, we have see this shift at UCLA with Google Apps for Education.

Learning analytics potentially provides access to finer-grained student data than has ever before been collected—often without student knowledge—within the walls or distributed beyond the walls of the institution to a variety of new partners and investors. It will be incumbent on universities to appreciate the breadth of the trade-offs and understand where the borders of existing policies meet the promises of cheap, ubiquitous technologies. Use cases will determine how learning analytics are designed, secured, implemented, and distributed: If the goal is individual student advising, then the data will remain personally identifiable and therefore protected; if the goal is measuring student usage of system, tools, or content for purchasing or other summative reasons, then depersonalized, anonymized data can be more widely distributed.

Reacting to the growth and speed with which data from online learning is emerging, in June 2014, the Asilomar Conference released six principles for discussion to guide online learning data; here we extract key features of these principles:

1. **Respect for the rights and dignity of learners.** Data collection, retention, use, and sharing practices must be made transparent to learners, and findings made publicly available, with essential protections for the privacy of individuals.

2. **Beneficence.** Individuals and organizations conducting learning research have an obligation to maximize possible benefits while minimizing possible harms.

---


21 [http://www.ihep.org](http://www.ihep.org)


3. **Justice.** Research practices and policies should enable the use of learning data in the service of providing benefit for all learners.

4. **Openness.** Whenever possible, individuals and organizations conducting learning research have an obligation to provide access to data, analytic techniques, and research results in the service of learning improvement and scientific progress.

5. **The humanity of learning.** Digital technologies can enhance, do not replace, and should never be allowed to erode the relationships that make learning a humane enterprise.

6. **Continuous consideration.** Ethically responsible learner research requires ongoing and broadly inclusive discussion of best practices and comparable standards among researchers, learners, and educational institutions.

Another set of concerns are the policies surrounding ownership of data. As data becomes more important for the learning enterprise, it demands connection to multiple sources that have been traditionally managed by separate entities. Certain high-level, cross-sector data communication is already common in a university, such as a counselor’s overview of a student’s academic profile from the Registrar or direct web links between content in the Library and the LMS. These links are fairly static in nature. Event-driven data, as described in the Khan Academy example earlier, requires that a student's KSA becomes part of the learning model, which entails synchronous access to detailed, individual student demographic and past performance; this raises a host of new challenges.

Finally, a general concern about data collection in a public university is the possibility that data may be subject to information requests through California Public Records Act ("CPRA") or federal Freedom of Information filings. Although many of the data elements are protected through FERPA or institutional privacy policies, the introduction of data from non-traditional sources, such as social media, complicates this concern to a greater degree. Working with learning analytics under IRB research guidelines may provide an additional level of protection at the cost of increased administrative complication.

**Projections and Opportunities**

It is common knowledge that education in American research universities is very different than education in private liberal arts colleges, state public universities, K-12, and differs even further from the types of education provided by for-profit online universities and corporate training. It is also well known that education is rapidly changing: continued public disinvestment and regulatory encroachment into higher education is increasingly being supported by private foundations and corporations with presumptive attitudes about solving the nation’s educational woes. Educational data in the form of learning analytics will continue to play an evidentiary role in shaping the future direction of teaching and learning across all sectors of education.

In the face of this shifting landscape, some higher education traditionalists object that learning analytics embodied in learning outcomes and rubrics, and specific learning measurements do not have a place in a university research education. This is a deeply-rooted cultural value: the UCLA Guide to Undergraduate Course and Program Approval, for example, supports a holistic view that, “A university course should set forth an integrated body of knowledge, with primary emphasis on presenting principles and theories rather than on developing skills and techniques.” Therefore the use of business intelligence tools that label students as “consumers” and courses as “inventory,” only reinforces these objections. One reason for this divide may be what former Harvard President Derek Bok identifies as differing goals of higher education: faculty “are occupied with the challenge of discovering and transmitting knowledge and ideas,” but often students, the general public, and business leaders view education as the primary entry point to careers.24

The following bullet points are intended to describe the potential for learning analytics in higher education, and to suggest that the market, federal and state governments, and accreditation are powerful forces that

---

threaten to trump academic tradition. Should these projections pan out—although some in the current higher education culture are wary of data “bean counters”—we may reach a point where this turns and we become wary of the naysayers.

- Although this paper is about data and learning analytics, at its heart, education and learning are fundamentally human endeavors, only fully realized when individuals take responsibility for their own scholastic advancement. Learning analytics will best succeed by supporting faculty and students and informing institutional decision making.
- Some people reject the idea of a data-driven education, but when data becomes information it proves reflectively self-worthy. This means that the value of the information gained from learning data will outweigh the resistance to seeing the value of data. We see this in the surprising accuracy of Amazon’s “people who bought this also bought” and Netflix’s “you may also like” recommendation services. These predictive tools are often accurate and make us painfully aware we aren’t as unique as we believe ourselves to be. Applied to education, this data promises to inform new discoveries about how people learn. Adam Cooper (JISC Centre for Educational Technology and Interoperability Standards) notes, “The volume of data available from a MOOC is large enough that questions can be answered that would be impossible to do so with an acceptable level of statistical significance with a traditional education cohort.”
- Automated, machine-driven learning can be expected to increase because it eases the strains of limited educational resources across all grade levels, pedagogical modes, geographies and constituents (local, state, federal and global).
- Evidence-based results will continue to be expected by economic and political forces.
- Online education presents a unique set of new data challenges surrounding instructional cost. Traditionally, educational cost accounting resides in the institutional or division level (cost per student, cost per degree). In a traditional lecture course, instructional costs may be limited to faculty and TA wages and benefits, and augmented by common classroom equipment costs. Costs for laboratory equipment are budgeted at the department level. Courses that require additional specialized elements or equipment handle costs through individual instructional improvement grants. Online education, however, represents a new variety of institutional per-course costs and therefore cost measures and accountability encroach into the digital classroom and into the practice of teaching in entirely new ways (such as captioning from fee-based transcription services). These courses and their costs will be aligned with programmatic need, marketing, enrollment and learning outcomes to ensure return on investment.
- Programmatic changes, like increased online courses that trigger the 50% WASC “substantive change” review or new online self-supporting programs, will turn to learning analytics to provide the evidence for learning-based ROI.
- The UCLA student of tomorrow is learning with tools like these today in K-12, so campus should be familiar with this kind of learning. Incoming students may expect learning methods similar to those that proved successful for them in K-12.
- Campus will need to provide new study skills to assist students adopt to the type of learning in higher education. That is, as machine-driven education increases in K-12, students need to adapt to learning the Socratic method, writing critical analysis papers, and gaining insight from other non-automated approaches.
- Analytic tools will play an increasingly important role in higher education. STEM disciplines are the obvious example, but other fields have been influenced by data and technology, such as the impact of genetics research on Biology and Franco Moretti’s data mining work on the canon of Western European literature.

---


26 It may be valuable to hear from faculty and academic counselors on the differences in student learning between the increasingly automated, simulation-based methods in K-12 and the traditional models of the university.

• In addition to impacting existing fields, it is not unreasonable to suggest new disciplines founded on manipulating data will emerge, such as Data Visualization and Probability. Although it is currently rooted primarily in Statistics, for example, Probability emerges most successfully from specialized measurement and data collection methods, data mining and data analytic algorithms, reporting, and scenario planning, etc. Probability is already directly impacting fields such as strategic planning, business, career counseling, law, pharmaceutical studies, transportation, and medical diagnosis.

• At an institution level, state authorization (or, state approval) continues to percolate. This Department of Education initiative will mandate that universities offering online programs meet the educational requirements in other states in order for remote students to earn credit. This means that educational artifacts of the programs, such as fully articulated learning outcomes and rubrics, will become requirements. These in turn become demonstrable data evidence for certificates or credentials. Although this approach to education is not generally known in research institutions, it behooves us to embrace the changes as soon as we can.

There is a portion of university education that will never rely on this type of analytic learning. Certain courses may never utilize systems such as the LMS that allow automated collection and processing of data. There are, however, courses that could utilize data, even where the possibility appears to be remote. For example, Humanities courses are often contrasted sharply with Computer Science courses across a wholly subjective and objective border. In many cases, this division of content and pedagogical approach is not clean. First, both disciplines—not just the Humanities courses—contain a wealth of knowledge that requires subjective judgment and critical thinking. Courses in algorithm design or network optimization, for example, rely on balancing a number of conflicting factors to determine the best approach to take. These lessons require considered trade-offs: there is no “one right way” but there are many completely wrong ways. Similarly, language-based courses, such as interpretation of poetry or translation tasks, rely on trade-offs between competing factors – it is safe to say there is no “one right way” to read a poem or translate a work, but there are many completely wrong ways.

On the other hand, even in courses where the final grade heavily favors a student’s ability to respond in a subjective manner, course work may include portions that can be objectively measured. Candace Thielle uses “unpacking the course” to label the process of identifying those portions of a course where automated technologies may be appropriately used, such as defining specific terms, identification, etc.

Data shifts the needle from anecdote and belief to rational, quantifiably defendable information. Some people will resist this but will be hard pressed to support their reluctance. Lessons from Google may soon eclipse the voices of protestors who argue we cannot measure learning:

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.28

In closing, two examples may inform the overall portrait of learning analytics and its potential for changing the landscape of higher education. First, describing the failure of data to influence a campus LMS platform decision, Macfayden and Dawson warn of the impact of existing culture resistance within a higher education institution:

An exclusive focus on technology integration issues, in the absence of development of a pedagogical vision, quickly neutralizes the likelihood that learning analytics data may catalyze organizational change with a focus on the student experience and learning outcomes. A focus on technological issues merely generates “urgency” around technical systems and integration concerns, and fails to address the complexities and challenges of institutional culture and change.29

Second, in May 2014, an email message from the UCLA Dean and Vice Provost encouraged faculty engaged in undergraduate instruction to participate in a survey as part of the National Science Foundation’s Widening Implementation & Demonstration of Evidence-based Reforms (WIDER) program. According to the email message, the WIDER program,

seeks to better understand the persistence of underrepresented minority students and women in certain fields of study. The survey results will provide a comprehensive profile of the teaching culture at UCLA, and asks questions about perceived institutional priorities, levels of satisfaction with and extent of teaching responsibilities, professional development experiences, instructional goals for students, learning strategies used in the classroom, and other useful information that pertains to teaching and mentoring undergraduates.  

In addition to the goals stated in Dean Turner’s message to faculty, the NSF program solicitation also contains a subtext related to both learning analytics and Macfayden and Dawson’s observation regarding “complexities and challenges of institutional culture and change.” The NSF’s program solicitation states:

The chief goal of WIDER is to transform institutions of higher education into supportive environments for STEM faculty members to substantially increase their use of evidence-based teaching and learning practices. …Through WIDER-supported projects, WIDER also seeks to learn if change theories and models from business, K-12, and the social sciences can be adapted to promote broadly within institutions evidence-based instructional and curricular improvements in undergraduate STEM by accomplishing the underlying cultural change that would often be required... . . . The most competitive implementation proposals will be based on explicit theories of change and will test hypotheses about transforming undergraduate teaching in STEM by examining the impact of deliberate processes of change... . . . Bold ideas for dramatically changing the status quo are especially sought. Although it has been observed that significant change in higher education is particularly hard to achieve... , the prospects for bold change in the delivery of effective higher education have never been better (Emphasis added).

Given the experience of LSU noted earlier, it is particularly interesting to note the NSF’s attempt to change university culture through its overt investment in evidentiary data and change theories and models from business and K-12. In light of opposing cultural and political factors like these, it will be fascinating to see the state of higher education twenty years hence.

---

30 Turner, Patricia. “Teaching Practices and a Chance at an X Parking Permit.” UCLA Undergraduate Education. 23 May 2014. E-mail.