The Predictive Learning Analytics Revolution

Leveraging Learning Data for Student Success

ECAR

ECAR Working Group Paper | October 7, 2015

ECAR working groups are where EDUCAUSE members come together to create solutions to today's problems and provide insight into higher education IT's tomorrow. Individuals at EDUCAUSE member institutions are invited to collaborate on projects that address core technology challenges and advance emerging technologies important to colleges and universities. More information can be found at the ECAR working groups website.

Introduction

Many might be surprised to learn that the United States, once a leader in educational attainment, is now ranked 12th internationally with regard to the number of 25- to 34-year olds who have an associate's degree or higher.[1] Given that we operate in a global economy and many jobs now require some level of higher education, this issue has risen to national prominence. Everyone from government agencies to the mainstream press are looking for and, increasingly, demanding improvements, particularly in areas that typically fall under the umbrella of "student success," which includes issues such as course completion, overall college retention, content mastery, and learning outcomes. As important as this is in the United States, the desire to improve student success is also a growing concern at institutions around the world.

It is in this environment that learning analytics has become a buzzword in higher education. Universities are seeing the potential that the vast quantities of data produced in their IT systems hold for addressing strategic challenges facing academia today. Whether analytics is used to identify ways to reduce the cost of education or to provide early interventions that help a struggling student succeed in a course, the combination of embedded administrative and academic technologies, big data, powerful analytical tools, and sophisticated data-mining techniques is poised to spark a revolution in how education is delivered—and in how the efficacy of that education is measured.

Such bold statements must be tempered with the recognition that the use of analytics in higher education is an emerging field. Parallels can be drawn with the learning management system (LMS) activity of the late 1990s. When LMSs began to be deployed at higher education institutions, many saw the strategic importance of the emerging technology but were unable to fully appreciate the role it would play in ushering in fundamentally new modes of delivering education. Similarly, while there is some consensus that analytics holds strategic importance for higher education, the specifics of the innovations that analytics will drive (or the failures that will occur) are largely speculative, pending further research to effectively inform practice. Nonetheless, it behooves institutional leaders such as the CIO and provost to gain an understanding of this emerging field as means to guide their institutions into a future where learning data play an increasingly important role in the process of education. This paper aims to enable institutional leaders and practitioners to educate themselves about the emerging field of predictive learning analytics and understand how it will impact the higher education landscape.[2]

EDUCAUSE

What Is Predictive Learning Analytics?

Analytics in higher education has many meanings and applications, ranging from the use of data to improve business operations—often referred to as "academic analytics"—to uses that more directly impact and assist both the learner and the learning process, generally referred to as "learning analytics."[3]

Learning analytics remains a relatively new term, and, as a result, several definitions exist in the literature today.[4] Although readers are encouraged to explore these sources and inform themselves of the varying approaches to defining the term, this paper uses what we feel is the most practical and concise definition:

> 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.[5]

This is not a new concept in higher education—almost since the dawn of modern teaching activities, higher education has been using data (e.g., quiz grades, course evaluations, etc.) to improve student learning. However, today's analytical systems have given us the ability to gather large volumes of data centrally and consistently, analyze them quickly, and distribute the results of analysis broadly in ways that are easy to understand and act on. Further, the development of sophisticated machine-learning data-mining techniques, as well as big data storage and processing capabilities, has allowed us to go beyond conventional reporting about the past and move into an era where we can predict, with reasonable accuracy, everything from future student learning outcomes (e.g., a student's final grade in a course) to whether a specific student will obtain a degree or continue in a given program. This ability to accurately predict future outcomes using learning data—called predictive learning analytics—is of significant strategic value because it empowers stakeholders in the learning process (e.g., students, faculty, administrators, et al.) with intelligence on which they can act as means to achieve more desirable final outcomes.

Key Terms

analytics The "discovery and communication of meaningful patterns in data."[1]

predictive analytics Encompasses "a variety of statistical techniques from modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future…events."[2]

learning analytics 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."[3]

predictive learning analytics The statistical analysis of historical and current data derived from learners and the learning process to create models that allow for predictions that improve the learning environment within which it occurs.

1. "Analytics," *Wikipedia*.
2. "Predictive analytics," *Wikipedia*.
3. 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, February 27–March 1, 2011.

As powerful as predictive learning analytics is, it is important for institutional decision makers to understand it as part of a larger process and that deploying sophisticated analytical tools and predictive models on their own will not likely impact student learning. In fact, the use of predictive learning analytics tends to come in the middle of the process rather than at the start or end (figure 1).

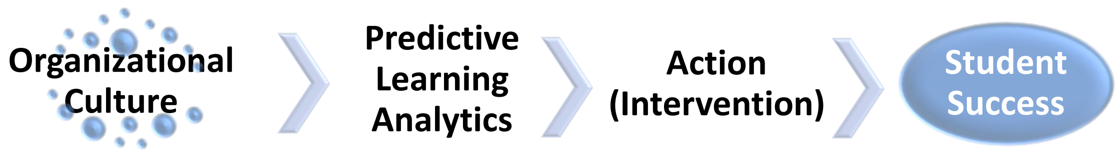


Figure 1. Position of predictive learning analytics in the path of student success

Before deploying predictive learning analytics solutions, an institution should ensure that its organizational culture understands and values data-informed decision-making processes. Equally important is that the organization be prepared with the policies, procedures, and skills needed to use the predictive learning analytics tools and be able to distill actionable intelligence from their use.

Once the necessary organizational culture has been established and predictive learning analytics systems have been deployed (see more on these steps in the sections that follow), action—often in the form of an intervention—needs to be taken. Analytics without action is merely reporting; interventions based on analytics are needed to improve student outcomes. Interventions may include setting policies, defining processes, or making referrals for academic or other support services. Interventions can be passive—such as a simple notification to a student or faculty member—or proactive, such as requiring a student to meet with an advisor. Interventions can be deployed by automated systems that require no direct human input or actively curated by student support professionals using integrated planning and advising for student success (iPASS) systems. Interventions should ideally have a specific, defined outcome and be measurable for effectiveness. When the impact of an analytics program or effort is measured in practical terms, it is almost always the interventions and their impact on outcomes being measured, not the quality of the tools or analytics themselves. Ultimately, with the right organizational culture, predictive learning analytics tools, and interventions, institutional leaders can expect to have a measurable, positive impact on overall student success metrics such as graduation rates, semester-to-semester persistence, and course-completion rates.

What Is an Intervention?

An intervention is any action that is taken with the intention of improving student outcomes. An intervention is ideally supported by analysis with measurable outcomes. Interventions can be passive or intrusive and can be facilitated by a person or be fully automated by a system.

This paper discusses how the results of predictive learning analytics are being used today to improve student success; shares key predictive learning data sources; highlights factors to consider when working with these data; and identifies strategic implementation considerations.

Predictive Learning Analytics Uses

The most important aspect of any analytics effort is the resulting ability to make decisions and take action on the data. In most educational settings, a range of stakeholders is involved in the teaching and learning process who can benefit from the results of predictive learning analytics. In this section we look at some of the primary stakeholders and provide practical examples of the types of predictive learning analytics results they are using, how they act on these results, and the impact their actions have on issues related to student success.

Students and Faculty

Students are often direct consumers of learning analytics, particularly through dashboards that support the development of self-regulated learning and insight into one's own learning.[6] An early example of this comes from Course Signals,[7] originally developed at Purdue University. As shown in figure 2, this dashboard uses symbolic traffic lights to warn students when they are at risk in a course (with a red signal) or to inform them that they are on track (a green signal).

In addition to predictive learning analytics that helps students at the course level, solutions are also emerging to assist students at the program level by predicting which students may not complete their degree on time or which courses would be best for a specific student to take next. One example is the learning analytics functionality embedded in the Blackboard Learn Ultra course view, which conducts background analysis of student LMS activity and grades in a course and provides automatic notifications, visualizations, and suggested actions within the LMS workflow (figure 3).

Figure 2. Course Signals

Another example is D2L Brightspace Degree Compass.[8] Originally developed at Austin Peay State University in Tennessee, Degree Compass is a recommendation system that assists students in course selection. Inspired by systems developed by companies such as Netflix, Amazon, and Pandora, Degree Compass pairs current students with the courses that best fit their talents and program of study for upcoming semesters. Degree Compass determines which courses are needed for the student to graduate and ranks them according to how they fit with the sequence of courses in the student's degree program and their centrality to the university curriculum as a whole. That ranking is then overlaid with a collaborative filtering model that predicts the courses in which the student is most likely to achieve the best grades.

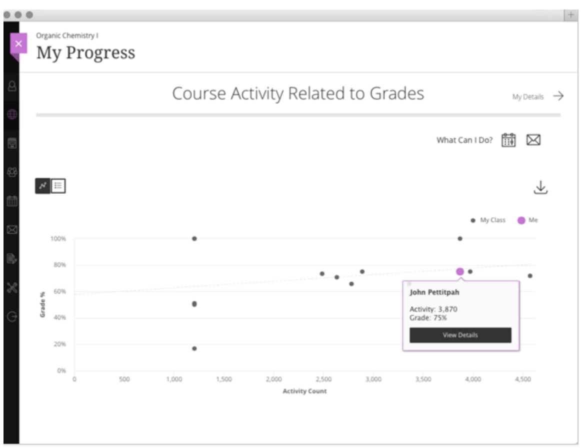


Figure 3. Blackboard Learn Ultra

Dashboards for faculty, advisors, and tutors support instructional staff by identifying trends and enabling early intervention. For instance, Brightspace LeaP by D2L works with course learning objectives, content, and questions and provides a text representation for each component. It then uses semantic algorithms to find relationships among these components to make intelligent recommendations for what should be presented to a learner to meet a particular learning objective, what questions should be used to determine if a learner has met the objective, and what content items the learner should read if a particular question is answered incorrectly. Instructors can further adjust the relationships to reflect their experience on how well certain content items have helped students in the past.

A system aimed at advisors is Student Explorer from the University of Michigan (figure 4).[9] Advisors use this system to identify the students who are at the highest risk of failure by providing a graphical view of achievement and a prioritized list of students for tutors to engage with.

Other applications let instructors see where students stack up against each other in a course using specific metrics—such as course access, content access, and social learning—overlaid with what those numbers typically mean for academic performance. For example, the Brightspace Student Success System (S3)[10] developed by D2L uses regression models that predict student grades starting from the first weeks. Instructors can monitor the status of individual students in terms of their predicted success. There are three levels of success, indicated by a color and shape: at risk (red triangle), potential risk (yellow diamond), and successful (green circle). The levels are determined based on thresholds on the predicted grade. The defaults are 0–59% for at risk, 60–79% for potential risk, and 80–100% for successful. These thresholds can be configured for each course during the setup of the predictive model.

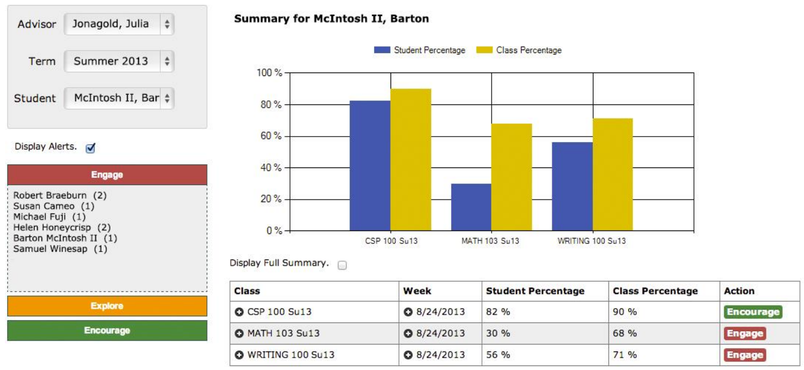


Figure 4. Student Explorer

S3 provides a risk-quadrant visualization that positions the student based on the success index—the overall predicted outcome, based primarily on engagement—and the student's current grade in the course (figure 5). The quadrant identifies each student in the class with a dot, with the current student highlighted. The quadrants are as follows:

1. **Withdrawal/Dropout Risk:** Identifies students who seem to be struggling in terms of both engagement and performance. In this case, the students may be at risk of dropping out.
2. **Academic Performance Risk:** Identifies students who seems to be engaged but are struggling in terms of performance.
3. **Under-Engagement Risk:** Includes students who appear to not be engaged yet are achieving high grades. In this case, the students may be under-challenged.
4. **On Track, Not At Risk:** Identifies students who are engaged and achieving high grades.

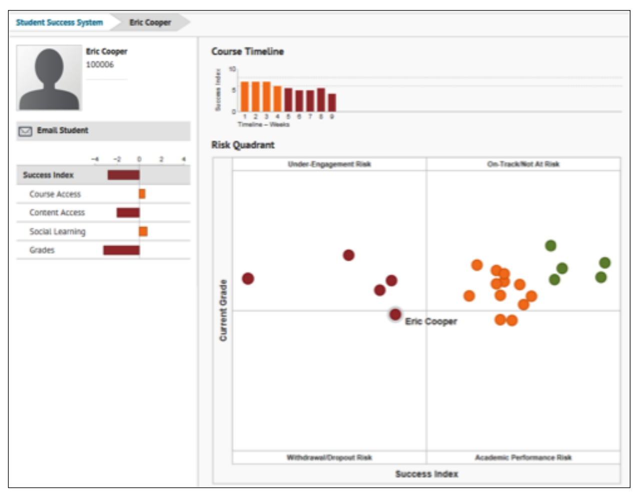


Figure 5. D2L Brightspace Student Success System

Another example is Blackboard Analytics for Learn, which combines data from the Blackboard Learn LMS with student demographic information and course attributes to create reports and dashboards for faculty, in order to provide a broad range of insight into course materials, student engagement, and student performance.

Advisors

Some institutions have begun to notify student advisors of a need to take action—often mandatory—based on the results of predictive learning analytics. For example, based on work with the Predictive Analytics Reporting (PAR) Framework (figure 6), the University of North Dakota (UND) is using analytics to drive policy and to guide operations.[11] In the policy area, findings from PAR around the impacts and predictive nature of withdrawals were used to change university culture on withdrawals so that to drop a course students are first asked to meet with an advisor, who can make sure that students are aware of tutoring opportunities or other services before withdrawing from the course. On an operational level, PAR risk scoring for individual students is loaded into a UND instance of the Starfish[12] system—an example of iPASS—to be used in multiple ways, including as an attribute of the overall risk score and to drive early alerts.

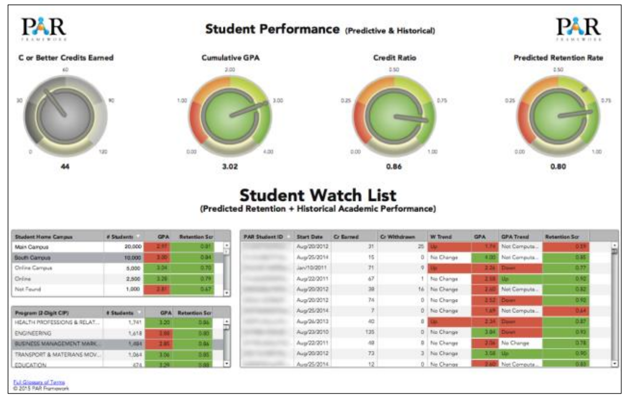


Figure 6. PAR Framework

Administrators

At the institutional level, administrators have long used predictive models for enrollment management, but the expansion of predictive modeling is now focusing on student success, completion, and operations. Simple examples are found in the use of predictive models to forecast the number of courses needed to meet student demand, based on predicted course-taking behavior. In more sophisticated uses, predictive models are used to forecast and understand the potential impacts of specific intervention strategies and the ROI potentials to help rank and make decisions about what efforts to focus on. These tools are fast becoming an evidence-based approach driving the use of scarce resources to improve student outcomes.

Predictive learning analytics is beginning to influence policy not only at the institution but also at the system and state levels. For instance, the Tennessee Board of Regents is working with all of the state’s publicly funded colleges to implement a “guided pathways” approach for their largest program areas, mapping out paths of study for students to follow in the hopes of improving college completion. This change in state-level higher education policy is a direct result of the predictive analytics work done at Austin Peay Community College with Degree Compass.

At the system and consortium levels, research using predictive modeling techniques is being applied, for instance, to students transferring from two-year colleges to four-year programs and trying to predict student success or support needs before students arrive on campus. PAR Framework has been working with multiple higher education systems to study the predictive attributes of students transferring from community colleges to bachelor’s-granting institutions to help shape curriculum, transfer pathways, and policy.

Much of the data for this analysis resides within the student information system (SIS) but requires aggregation and transformation to provide these insights. For example, the Blackboard Analytics Student Module helps institutions improve retention through targeted analysis that cross-references students’ demographic attributes with their performance. Based on this analysis, institutions and their advisors can influence specific student groups’ performance through targeted student programs.

Key Predictive Learning Data Systems and Sources

Traditionally, institutions have focused on sources with static or historical data (e.g., student demographics, grade point average, residency status, etc.) to identify students who may be at risk of low achievement and who may benefit from additional support services. While these data sources are statistically important predictors that remain relevant in the field of learning analytics, we can also tap more dynamic data produced from a range of instructional technologies (such as LMS event log data, electronic gradebook data, attendance data, library data, etc.). These dynamic measures of student learning, effort, and engagement—when combined with traditional measures—allow for a more nuanced and personalized analysis. The significance to higher education is that student success (or failure) can be predicted with more accuracy, earlier in the learning process, than ever before. Such insights, in turn, are allowing institutions to intervene much earlier in the academic term or, in the case of learning activities, before the student moves on to the next assignment. Early interventions of this nature can dramatically improve the odds that students will receive the help and support they need to be successful before it is too late.

When contextualized within course-level pedagogical frameworks, predictive learning analytics can be used to forecast complex patterns of student learning by tracking student behaviors within an LMS (e.g., quiz performance, engagement in online discussions, or online collaborative work). These patterns can subsequently be used as the basis for real-time interventions with students (e.g., automated advice on links to materials/hints or suggestions on learning approaches).

Open Academic Analytics Initiative

Supported by the EDUCAUSE Next Generation Learning Challenges program and funded through the Bill & Melinda Gates Foundation, the Open Academic Analytics Initiative (OAAI) has developed and deployed an open-source academic early-alert system that can predict (with 70–80% accuracy) within the first two or three weeks of a semester which students in a course are unlikely to complete the course successfully. Once at-risk students have been identified, they can receive interventions, such as tutoring or other resources, designed to help them succeed. Research has shown that use of the OAAI system had a statistically significant impact on final course grades and content mastery. See "Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative."

As institutions begin to plan for and implement learning analytics, it is important to develop an understanding of how these data are collected across the institution and can be used to improve student success. The following sections look at the three most frequent kinds of predictive learning data that exist in higher education: activity data, achievement data, and static data.

Activity Data

One of the most frequently used sources of activity data is the LMS, which provides vast amounts of data about learner interactions with content, assessments, subject-matter experts, and peers. The LMS has become a mainstream technology for both online and face-to-face courses, in place at 99% of higher education institutions.[13] It is worthwhile noting, however, that use of the LMS varies a great deal among faculty and

Key Strategic Activity Data Questions

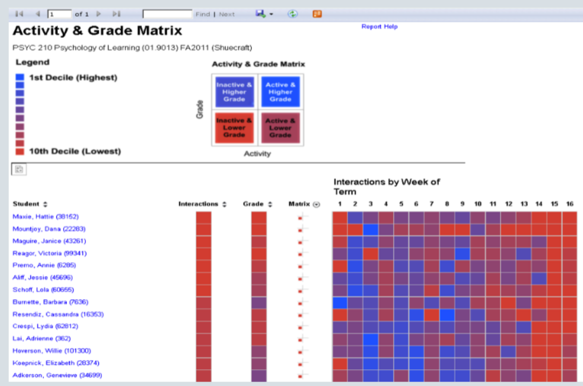
- How long does the institution retain LMS event log data?
- What percentage of courses use the LMS? What percentage of faculty and students use the LMS?
- Is there deep integration of the LMS in student learning, whether in face-to-face or online courses?
- What educational technology tools does the campus use that do not make log data available?

programs, ranging from "power users" to those who may only post a syllabus and a few files. The recent ECAR study, *The Current Ecosystem of Learning Management Systems in Higher Education: Student, Faculty, and IT Perspectives*, notes that "the ways in which [faculty] typically use the LMS are less about interaction or engagement activities and more about sharing content with students." Additionally, only 56% of students indicate that they use the LMS in "most or all of their classes."[14]

Most learning analytics systems are capable of providing results even when limited use of LMS tools results in "missing data." In general, however, the more the LMS is used (and thus the more data are produced), the better the analytics results tend to be. One way to mitigate this issue—beyond requiring more LMS use—is to provide users with a "confidence rating" that helps them understand the accuracy of the prediction. For example, the rating may be lower if less data are available for a specific student or course.

Heat Maps of Course Activity

Application providers are providing visualizations that transform activity logs and other LMS-captured data into information that enables instructors to quickly identify patterns and exceptions to standard behavior, thereby helping them better understand their students' study activities. The screenshot below from Blackboard's Activity for Learn application charts activity and grade for each week of the course.



Regardless, the widespread availability of LMSs in higher education and the diverse set of instructional tools it provides make the LMS a convenient and important source for rich learning data. The two primary LMS data sets that are most often used for predictive learning analytics are activity data from event logs (discussed here) and achievement data from electronic gradebooks (discussed in the next section).

Clickstream Activity Data

Although building predictive models out of clickstream or activity data—data related to what users click on as they navigate a system or course—is in its infancy, a growing number of researchers aim to do this with the hope of building more responsive predictive models. For instance, researchers at the University of Michigan are building daily clickstream logs from the Coursera MOOC platform, allowing for the prediction of learner outcomes with increasing accuracy over the semester.[1]

1. Christopher Brooks, Craig Thompson, and Stephanie Teasley, "A Time Series Interaction Analysis Method for Building Predictive Models of Learners Using Log Data," in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, ACM, March 2015, 126–135).

Event log data are collected as users interact with the LMS. Event logs tend to include data at different granularities, from high-level general events, such as number of course-site logins, to fine-grained information, such as the number of individual content views or engagements with discussion forums. Whereas a series of high-level events may be easily shown to an instructor as a table or other simple visualization, fine-grained events might require significant programming and machine-learning skills for an analyst to understand.

This work can be further complicated by the fact that because event log data were originally collected as a means to monitor server performance and troubleshoot technical issues, a significant amount of effort is often required to extract large volumes of data and complete the analytical processing required to use these data for learning analytics. In some cases, institutions may even find that the event log data they need have not been kept for any significant length of time, which poses a barrier to leveraging historical data to create predictive models. In addition, the format of log data is highly variable, both between and within educational technology systems, which can add significant overheard in the data sense-making process. Despite the challenges working with these types of data, the end result can provide tremendous benefit to students, faculty, and the institution, particularly with respect to improving overall student success.

Although the LMS is the most common source of activity data, other sources have shown to be beneficial as well:

- **Digital Content:** Publishers, both open source and commercial, have long used educational research and analytics about student interactions with their materials to inform development. In recent years, publishers have begun to offer analytics dashboards and reports that provide insights into student time on task, completion of course activities, and predictions of students struggling with course materials. Most of this functionality resides within the publisher application, but data feeds might also be available to supply a predictive analytics engine with data from multiple sources.
- **Personal Response Systems (clickers):** These systems capture attendance and can track other class responses at in-person lectures and provide formative assessment of student content mastery.
- **Video-Streaming Servers:** With increasing detail, video-streaming servers demonstrate student interaction with videos, including basic time on task and more detailed interactions (e.g., when a student rewinds, fast-forwards, etc.).
- **Web Conferencing:** In this medium, participation in virtual lectures, archived sessions, and other online activities can be captured.

Achievement Data

The LMS also provides a rich set of electronic gradebook achievement data. This tool generally includes grades associated with examinations, assignments, and participation (e.g., in online discussion). Such grades are of value when trying to analyze student success in a specific course associated with specific learning activities. Because most LMS gradebooks were not designed with learning analytics in mind, extracting and analyzing the data associated with them can be nontrivial. For example, understanding how a student did on a particular quiz without also understanding the weighting that quiz has on overall grades can be misleading. Thus, it is necessary not only to extract individual grades but also to contextualize these grades, such as through weighting and retry opportunities.

Combining Survey, Student Aptitude, and Demographic Data

The University of Saskatchewan has begun to build predictive models that include aptitude, demographic, metacognitive, and academic achievement variables for certain freshman courses. Data are aggregated from SISs, LMSs, an entrance survey, and usage of learning resource centers. The entrance survey contains rich demographic and learning-related items, such as motivation, grit, and learning strategies, and has high response rates (55–60%).

In 2014, weekly, personalized advice messages (based on expected grades and demographic variables) were delivered to each of the 1,200 students in a one-semester freshman biology class. Advice templates were crafted by learning specialists to encourage students to use learning resources and supports appropriate to their expected individual needs. At the end of the course, using the same evaluation rubric as in prior years, there were significantly fewer withdrawals, significantly fewer D’s and F’s (30% fewer) and significantly more A’s (23% more) than in the two previous years.

For more information, see The Student Advice Recommender Agent: SARA.

Static Data

Since well before learning analytics came on the scene, higher education has been using static learning data—data that do not typically change during a course—to predict how students might perform toward their academic goals (e.g., scores on the SAT and other standardized tests such as placement exams). Although the increasing availability of dynamic learning activity data has improved the accuracy of such predictions, the inclusion of more traditional static learner data remains important, given the longstanding correlations that have been established between these data and student success.

Whereas a range of static data are used in learning analytics work, much of it falls under the broad umbrella of student aptitude and demographic data (e.g., high school GPA, race/ethnicity, or academic standing) that may in rare cases change over time but generally remain constant. In most cases, student aptitude and demographic data reside in the institution's SIS and are archived for reasonably long periods of time. This can make the data-modeling process easier and allow for the development of predictive models based on large amounts of historical data. At the same time, the data associated with such systems are often considered confidential, protected by laws such as the Family Educational Rights and Privacy Act (FERPA).[15] As a result, accessing and using these data can take time to obtain the necessarily approvals and will often require that rigorous data-security protocols are followed.

Current learning analytics research suggests that other static data sets—although less frequently used (or collected)—may be of value in predicting student success:

- **Survey Data:** Surveys can provide insights into student attitudes, perceptions, and fears, as well as goals, emerging interests, etc.
- **Extracurricular Activities:** Data associated with clubs, sports teams, fraternities/sororities, and other activities outside mainstream coursework can help determine levels of student engagement with the college or university. This engagement level can be a predictor of things such as persistence to graduation.

Each form of data has its own data-access and data-storage challenges and may not be available at a given institution.

Learning Analytics Solutions

With learning analytics, institutions will need tools to collect, store, analyze, and visualize the data in meaningful and intuitive dashboards. Understanding the available solutions and what value each type provides allows institutions to understand how to best craft a solution to meet their unique needs.

There are two major categories of predictive learning analytics solutions in the market today: embedded and platform. There are two types of embedded solutions: LMSs that contain embedded analytics tools for use by existing LMS users, and SISs with embedded analytics tools that have built-in triggers or alerts that execute based on transactions or the lack thereof. Neither of these embedded solutions is by design broad enough in scope to give an institution the complete spectrum of data that represents a student's experience. In addition, these types of solutions do not have the capacity to accept and integrate/store additional data streams (e.g., from third-party tools that create additional data streams in the cloud).

Dedicated analytics platform solutions are focused on specific areas, such as at-risk student retention. These solutions leverage data streams extracted from a variety of traditional institutional systems (e.g., the registrar, LMS, and student information and administrative systems), which are then provided to advisors or other stakeholders for use in their work with students.

Once an institution has determined what *type* of solution best meets its needs, it must determine how to *implement* the solution. A college or university might want to build its own solution using internal resources, implement an existing solution (commonly a proprietary solution), or collaborate with others (often through a consortium) to co-develop a solution (which might be released under an open license).

Choosing a Solution: Critical Questions

- Will it integrate with other systems?
- What standards does the vendor use?
- What tools are available to help with the collection, description, and analysis of data in the system?
- How can data be retrieved and shared with complementary systems?
- How is historical data moved from one system to another?
- Will we need to change systems in the future?
- Is the vendor committed to the IMS Global standards and community?

- **Build Your Own**: Building a solution from scratch has appeal. It can be designed to meet specific needs and address the identified purpose; users can be certain about how the system operates; and the components of the system can be tweaked and swapped out over time. However, obtaining resources and either training or hiring staff with the requisite experience may be challenging and costly, as are the ongoing maintenance and support challenges of a custom, homegrown approach. Being in the driver's seat provides full flexibility but may ultimately be more costly than a purchased solution.
- **Implement an Existing Solution**: Selecting a complete solution—typically through purchasing a vendor-provided system—may allow for a more rapid implementation. As with any solution, however, additional costs and resources may need to be devoted to extract data and integrate the solution with other institutional systems. Special attention will also need to be paid to whether there are any limitations on how the system can be used and whether modifications can be made (and at what cost). The institution will need to determine if it has all the necessary features for analytics needs and understand the factors regarding data ownership, use, and curation.
- **Collaborate**: Collaborative efforts are emerging in two forms. Some institutions are working together to build analytics ecosystems, leveraging available open-source software components, while others are forming consortia to tackle these efforts together. Both of these collaborative efforts typically focus on implementing community specifications, such as those created by IMS Global (see "Specifications for Capturing and Storing Learning Activity Data" for more information). Open source lets institutions leverage the work of others and not bear all of the development costs alone. Shared governance means institutions can have a say in the future of the toolset, but lack of a vendor often means execution times depend on peers, the consortia team, or others in the community. Hybrid, vendor-supported, open-source systems are another option, providing many of the benefits of both approaches (but possibly also removing some of the cost benefits that a full open-source solution may offer).

A decision should take into account the organization's strengths to determine where the greatest positive impacts on the overall analytics ecosystem can be achieved. For example, if an institution has strong data skills but weak analysis capabilities, it may want to consider building local support for the data component and purchasing the analysis tools. Careful consideration is needed to select a solution that serves the needs of the institution and ensure it is poised for growth while controlling costs.

Working with Learning Data

A number of tools, methods, and standards are used to capture, store, extract, and analyze learning data within the context of predictive learning analytics. Although the details of machine learning and related data-mining techniques are beyond the scope of this paper, it is helpful for institutional leaders to have a high-level understanding of the general process that is used when engaging in this type of work. Figure 7 provides an overview of this process, which generally starts with extracting large amounts of historical data and preparing them for analysis. Once the data are prepared, data mining identifies patterns that correlate system usage with student success outcomes (e.g., final grades in a course, graduating on time, etc.). These patterns are then used as the basis for a predictive model and its underlying algorithms that analyze current students' data in near-real time. This predictive model analysis ultimately produces actionable intelligence, such as an alert to a faculty member that a student may be at risk of failing.

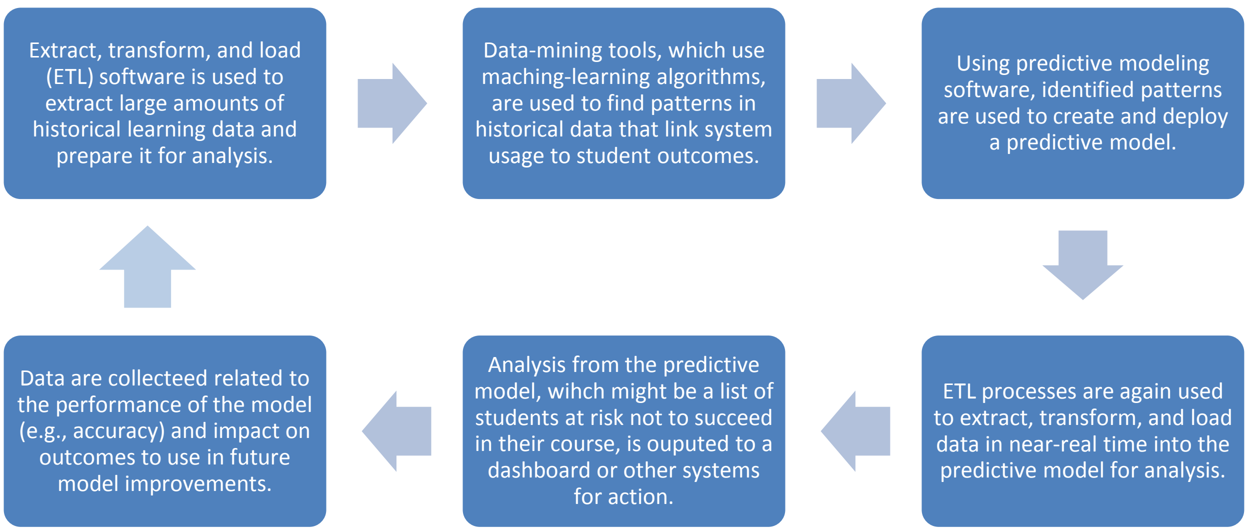


Figure 7. Overview of predictive model development and deployment

Many of the mechanisms for building predictive models that are specific to learning analytics are just emerging, and questions remain regarding which tools, techniques, or methods will be most prevalent in the future. This section discusses some of the current approaches to learning-data collection and processing and some of the emerging trends that will be important to watch.

Specifications for Capturing and Storing Learning Activity Data

Until recently, collecting learning data was done through fairly traditional methods of querying data out of system databases and transferring them to a common repository where they would often need to be "linked" or connected with data from other sources. For example, a database administrator might extract student demographic and aptitude data from the institution's SIS and event log data from the LMS and then bring both sets of data into a common database. In doing so, individual user event log data would be connected with the same user's demographic and aptitude data. Although this approach remains common today, particularly when working with static data such as student demographics, a number of emerging specifications are designed to collect learning activity or clickstream data in real time and route them into a common repository. Two organizations in particular—the Advanced Distributed Learning (ADL) Initiative[16] and the IMS Global Learning Consortium[17]—are providing specifications to describe interaction data as they are generated in learning environments.

Both the Experience API (xAPI)[18] from ADL and the IMS Caliper Analytics specification[19] are founded on principles from semantic web technologies[20] and are designed to allow institutions to capture learning activity data in a systematic way across multiple systems and platforms. Once captured in a standardized format and syntax, these data are sent to and stored in a repository using an emerging class of applications called the learner record store (LRS). The LRS serves as an endpoint for data collection from all of the learning technologies offered by the institution and describes the learners and their activities within these technologies (e.g., the LMS, clicker data, etc.). Although conceptually the LRS is considered a centralized service, several LRS solutions might be used in coordination. An open question is whether real-world deployments will focus on a single centralized LRS for reporting and predictive modeling services, or individual learning technologies will have unique LRSs, or both.

The LRS is similar to work many institutions are already doing with data warehouses and operational data stores that pull together all the information available about a specific student. Keeping an eye on the emerging LRS common data definitions[21] and being able to adhere to them when needed might well be a strategic advantage to an institution. It may be necessary to start with a limited range of data sources, but collecting as much data as possible into a common format will likely yield advantages in systems interoperability, benchmarking, longitudinal research, and data readiness for future analytics activity. Nonetheless, while the concept of LRSs has begun to gain acceptance in the educational technology space, it is unclear exactly how institutions will choose to assemble analytics warehousing from among the variety of products and services that are created from this movement.

As with any emerging standards, the xAPI, IMS Caliper, and LRS standards will need more time to mature and become established before their full strategic importance and value are known. Institutional leaders may want to pilot and explore these standards while watching this space closely over the next one to two years.

Issues with Data Access

With the move to cloud-based services, it's increasingly important that IT and other institutional leaders consider data-access, lock-in, and interoperability issues when evaluating the cost savings and other benefits of the cloud.[22] Whereas just a few years ago assets such as event log data were only needed on occasion for troubleshooting technical problems, data produced by institutional systems are now critical fuel for predictive learning analytics. Ensuring that all of these data can be not only accessed when needed but also migrated between systems is now a strategic imperative. Institutions should avoid situations in which they need to move to a new LMS but cannot because they will lose access to historical data or because the new solution will not allow data to be exported into a learning analytics platform. Similarly, be aware of costs that may be associated with accessing and extracting data from cloud-based services.[23]

Although data-access issues associated with cloud-based services is a timely topic, many IT systems today continue to be hosted locally by the institution, and thus it is equally important to understand the degree to which local data can be easily accessed, extracted, and shared. Learning data are critical to the deployment of almost any learning analytics solutions. As recently as a few years ago, it was fairly rare that anyone would need to access, let alone extract, these data other than for troubleshooting technical issues. Thus, determining whether data structures are documented and knowing whether the data are encrypted or if access to the data is allowed were not strategic issues. Due to learning analytics and its thirst for real-time and historical learning data that might span years, understanding how open the data are in these systems is a critical issue that every institutional IT leader needs to consider.

It is important to understand the availability not only of the data but also of the models behind the data. Predictive models will increasingly play larger roles in the decisions being made on campuses. Whether it is determining which students are in need of academic assistance or which courses a student should consider taking, decisions facilitated by these models will have significant and long-term impacts on learners and the institution itself. Today, the vast majority of the predictive models in higher education are closed—that is, they are licensed so that institutions cannot see the algorithms and logic being used. Inevitably, the time will come when a student (or parent) wants to know why the student was told he would likely fail a class, or a provost will expect an explanation for why a specific set of students is showing up on an alert dashboard. Possibly even more importantly, closed models might prevent institutions from adjusting the model to address local needs—such as a nontraditional program for which an out-of-the-box model doesn't work well—or to improve performance by integrating additional historical data.

Open-Source Data Extraction: Apereo Learning Analytics Initiative

Building on its EDUCAUSE Next Generation Learning Challenge project, Marist College partnered with the University of Amsterdam and several other institutions and organizations in 2013 to form the Apereo Learning Analytics Initiative, with the goal of creating an ecosystem of open-source tools to support predictive learning analytics.

In addition to the Learning Analytics Processor—a tool to extract data from an LMS, pass it through a predictive model to produce risk scores, and then display the data back to the LMS using the LTI standard—members of this initiative have developed and released an initial version of an open learning record store (OpenLRS), dashboard (Open Dashboard), and student intervention system (Student Success Plan). They are also building a library of open predictive models.

Being aware of the openness of learning data in these systems—and of the extent to which the predictive models can be viewed and changed—will increasingly become critical. Institutional leaders working in the learning analytics space may find it useful to make such questions standard operating procedure when deploying new systems or moving services to the cloud.

Processing Learning Data

Once data are brought into a common repository, they need to be processed for data mining and scoring (the process by which the predictive model algorithm generates a numerical score—generally a probability—for each student). This processing is accomplished through extract, transform, and load (ETL) processes, for which a range of software tools is available (e.g., SPSS and Kettle). For example, when working with LMS event log data—which first needs to be extracted, either via a database query or by using a standard such as xAPI—it is critical that individual student data be compared to course averages rather than simply looking at absolute values. Knowing that a student accessed a course 10 times in one week does not indicate whether this was a lot or a little without also knowing that the class average was 50 accesses per week. This type of data-transformation process is often time-consuming to set up, but once it is developed it can be reused over and over again to perform the same transformations. Once the transformation work is completed, the resulting data are generally loaded into the predictive application for analysis to complete the ETL process. Because raw event log and system data almost always need to be processed prior to being used in predictive learning analytics—sometimes extensively—so that they will be in a format that a predictive model can consume, developing an ETL process is a critical and necessary step. Institutional leaders should take note that this work is often overlooked or underestimated in terms of complexity, resource usage, and overall value to the success of the project.

Data Integrity

Quality data are the foundation of any predictive learning analytics initiative. Practices that ensure data quality need significant reflection. Most institutions have strict data-quality standards for some portions of their operations; however, that level of quality is difficult to sustain, and until recently, certain data sources did not require scrutiny. Pipino, Lee, and Wang[24] discuss 16 dimensions of data quality that offer institutions a framework for operationalizing data-quality audits on sources that do not already have them. The authors posit that all 16 dimensions must be systemically addressed to ensure foundational stability for any learning analytics endeavor. Institutional leaders and practitioners who deal with issues of data quality would benefit from reviewing these dimensions in more detail.

De-Identification of Learning Data

The Open Academic Analytics Initiative (OAAI) faced the challenge of needing to collect learning data from several institutions for analysis through its academic early-alert system and then return results to individual institutions and instructors. To address privacy and data-security concerns, a protocol was developed by which student identifying information, such as campus-wide ID and e-mail address, was removed from the original data sets and replaced with a randomly generated unique identifier. This was accomplished by using a "master key" to link real identities with the randomly generated code. Once identities were removed, all but one copy of the key were destroyed, with the only remaining key being encrypted and then held by each institution's institutional research and planning office. This allowed for analysis of the learning data without the possibility of identifying individual students. Only after results were returned to the instructors for follow-up were student identities again revealed.

Data Privacy, Security, and Ethics

Data privacy and security are not new concerns for higher education, but the use of predictive learning analytics and the data collection that accompanies it require the collection and analysis of unprecedentedly large data sets stored in centralized systems that contain confidential information about individual students. Although regulations vary between countries, in the United States much of these data are protected by federal laws such as FERPA, which carries significant penalties for institutions found to be in violation. In addition, new systems in this space (such as iPASS) now collect information not previously stored in institutional systems, such as food-stamp eligibility or other socioeconomic factors. Currently, FERPA is undergoing review at the federal level to increase transparency and restrict access to data, in consideration of new approaches to and the rising importance of data.

In general, two conceptual areas are important to understand when working with sensitive data and mitigating risk: data privacy and security. Privacy issues related to working with learning data are often effectively addressed through the de-identification of the data prior to use by researchers, technical staff, or administrators. This process involves removing any personal identifiers (e.g., name, e-mail address, etc.) from the data sets and, although not always necessary, replacing the unique identifier with a randomly generated code. Maintaining a unique identifier enables joining disparate data sets and analyzing individuals over time.

Securing these data, particularly any data keys that could reveal the actual identities, is equally important, given the serious legal and public relations consequences associated with data breaches. To this point, implementing proper data security protocols can often be just as important as the larger institutional community's understanding that the issue is taken seriously and that precautions are in place. Failure to communicate these points can lead to a *perception* of data security concerns that can often be as difficult to deal with as actual security problems. Data security is a well-established domain that is not unique to predictive learning analytics, and institutional leaders are encouraged to review the literature on this topic.[25]

Finally, the use of predictive learning analytics in the academy also raises a number of important ethical questions, such as whether the institution has an obligation to use learning data and act when it believes it has an opportunity to help students succeed. Another example is whether it would be ethical for a faculty member to limit support to students based on expectations of failure (e.g., not responding to e-mails from students who aren't actively participating in class). This remains an actively debated topic representing a range of opinions, so institutional decision makers should also spend time reviewing current literature related to ethics and learning analytics and engaging in the dialogue.

Data privacy and related ethical issues are often easily pushed to the side when institutions begin to deploy learning analytics solutions, as other challenges often appear more significant. Leaders ignore these topics at their own peril, however, because it takes only one incident, real or perceived, to make security a major topic on campus or even in the national news. Taking the time to understand the issues and address them up front can often help avoid a crisis later.

Strategic Implementation Considerations

Implementing a predictive learning analytics initiative requires that a number of decisions be made. Before getting started, consider these questions:

- What is the organization's existing capacity to undertake an analytics effort, or—if the current capacity isn't sufficient—how will the institution build capacity?[26]
- What type of predictive learning analytics solution will work best for the campus environment—an embedded or a platform solution?
- How will the institution deploy the solution? Will it build a solution, purchase an existing one, or collaborate with a consortium or otherwise develop a joint solution?

Key to implementation success is the human and organizational capacity to support a predictive learning analytics initiative. Building institutional capacity, regardless of whether an institution chooses to grow capacity internally or through partnerships, is essential in moving learning analytics from small-scale, research-based activities to scalable, systemic efforts. Strategic decisions will need to be made about institutional goals, readiness for learning analytics, desired scale, timeline, and resources.

In their 2013 publication *Building Organizational Capacity for Analytics*,[27] Norris and Baer adapt a business analytics framework and apply it to analytics for student success. The resulting model shows the necessary progression from standard reporting to analytics efforts. The institutional capacity necessary to implement analytics for student success cannot be discounted. Norris and Baer focus on five domains that affect capacity[28] and that serve as a framework for consideration. The domains, refined and adjusted to apply to predictive learning analytics, include:

- Leadership
- Culture and Behavior
- Technology Infrastructure, Tools, and Applications
- Policies and Compliance
- Predictive Learning Analytics Skills

In addition to using these domains as a framework, it is important that institutions view capacity contextually (i.e., what is sufficient for capacity at institution A may not be so for institution B) and systematically (i.e., with deliberate intention toward enterprise adoption). In addition, institutions must carefully consider which capacities are strategically important to have in-house versus outsourced. Another consideration is more pragmatic—whereas most institutions can easily find the *existence* of necessary capacity (especially in values and skills) somewhere in the institution, *access* to capacity is often limited. Finally, institutional reflection and established goals will be of paramount importance in formative stages as the march toward optimization is undertaken.

Leadership

Learning data are often spread across numerous systems controlled and overseen by a range of areas, from informational technology to deans to admissions offices. Often critical to implementing a strong institutional analytics approach is having a senior-level champion (e.g., a vice provost or vice president) who can help remove organizational barriers and build bridges to the various stakeholders and their data. In addition, departmental leaders need to take leadership in the coordination of analytics efforts—both within their own departments and with other departments—and faculty champions serve a valuable role as demonstrators of new approaches to learning.

Culture and Behavior

Broad adoption of learning analytics will rely on an institutional culture of continuous improvement and data-informed decisions. Predictive learning analytics enables data-informed decision making, and institutions that manage themselves by this paradigm will find these analytics a natural fit to their institutional planning processes. A common first step is familiarizing stakeholders with predictive learning analytics generally, the benefits and capabilities that analytics offer, and the expectations and goals of the institution’s use of analytics. Training and additional resources may also need to be provided to educate users on local practices such as the use of data (ownership, use, access, privacy),[29] to clarify the impact an analytics program may have on various stakeholders, and to provide opportunities for stakeholders to learn about and try different tools and processes.

Technology Infrastructure, Tools, and Applications

One of the foundational needs of any predictive learning analytics effort is the underlying technology. Whereas institutional IT has grown immensely over the past 20 years, the capacity for learning analytics is often not completely met by existing architecture and infrastructure. In particular, the architecture and infrastructure may not be robust enough to efficiently analyze the volumes of big data involved or may not provide sufficiently sophisticated analytical tools.

To better understand these challenges as they relate to predictive learning analytics, consider the explosion of educational technologies in higher education. A 2014 study at the University of Texas at Tyler indicates, "[T]he average faculty member utilized about six technology tools in their courses."[30] Common tools include the LMS, e-portfolios, presentation software, plagiarism detection, blogs, e-texts, video and capturing services, social media, web-conference services, and lesson-planning tools. All of these tools create clickstream data about learners in different ways and, although most institutions are able to leverage structured data coming from these applications, much of the data being captured in today's environment are unstructured and messy, making it very labor intensive to bring disparate data sources together in a cohesive repository that allows institutions to successfully weave together a more holistic view of the learning environment.

Practical issues for consideration in this domain include whether the institution has:

- Capacity to store or access disparate data sources in raw or transformed form
- Capacity to store or access predictive analytics results
- Capacity to deploy and measure the effects of learning interventions
- Capacity to integrate numerous predictive analytics tools
- Computing power for regular big data analyses, simulations, visualizations, and processes
- Security protocols in place that ensure the learning analytics effort is not a liability

Policies and Compliance

Predictive learning analytics is predicated on the assumption that data, including historical data, are available for analysis. As such, appropriate institutional policies and processes regarding data access and use are central components for successful implementations. In the United Kingdom in late 2014, the Open University adopted its "Policy on Ethical Use of Student Data for Learning Analytics."[31] Jisc, positing that "current legal and ethical guidelines have not caught up with innovations in the identification of patterns and new knowledge emerging from the vast datasets being accumulated by institutions," released a literature review of ethical and legal issues, titled "Code of Practice for Learning Analytics,"[32] followed by an actual code of practice in 2015.[33] At the highest level, policy specific to learning analytics has been ratified at surprisingly few U.S. institutions.

In the absence of policy specific to learning analytics, existing policy often is applied, or de facto policies may fall into place. As adoption of predictive learning analytics grows, however, systemic policies and practices will need to be developed, and it is expected that additional compliance may be necessary as federal and state laws focus more on this space.

Federal Laws

Currently, two federal laws commonly come into play when working with predictive learning analytics, FERPA and Section 508 of the Rehabilitation Act. FERPA, which was signed into U.S. law in 1974, protects the privacy of student education records. Section 508 was introduced in 1998 “to require Federal agencies [including higher education institutions that receive funding through the Assistive Technology Act] to make their electronic and information technology (EIT) accessible to people with disabilities.”[34] If a learning environment is not accessible, it may fail to record learning analytics data for students with disabilities, causing those students to fall out of the analytics scope and its reports, potentially being left out and left behind, resulting in an uneven playing field and hence a serious ADA issue. This would be the case even if the system recorded some, but not all, of the relevant learning data for impaired students.

Both FERPA and Section 508 were created before widespread adoption of the World Wide Web and the current profusion of educational data. Whereas both laws have been the topic of major legal conversations for quite some time, both still stand as they did over a decade ago. Work has been coming out of the Department of Education and the White House that attempts to reinterpret FERPA and other federal regulations in the context of big data solutions.[35] As these conversations progress, it will be important for institutional leaders to engage with regulatory bodies to ensure that student and staff protections are safeguarded while also enabling institutions to safely and effectively carry out analytics.

State and Local Laws

Increasingly, state and local laws focus on data protection, including student data. That said, each state is unique in its approach, and any institution engaged in learning analytics should carefully review local laws regarding educational data. For example, institutions should consider how a student’s privacy might be protected in the case of a request made under the state’s freedom-of-information laws, particularly if the individual has been flagged by predictive learning analytics as an at-risk student.

Institutional Policies

The importance of data governance and institutional policies that address data collection, security, retention, and use has gained a great deal of attention in the past decade, as both big data and analytics have come to the forefront.[36] In addition, more attention has been devoted to the establishment of formal data stewardship programs that define stewards’ roles and responsibilities, where stewards reside in an organization, and how they work with colleagues to ensure that data are maintained to meet both internal university responsibilities and external best practices.[37] Each institution will have unique policies that govern educational data usage and that may also vary based on data type. When establishing an institutional predictive learning analytics program, existing policies, processes, and workflows will need to be consulted to determine what restrictions may currently be in place and what guidelines might be needed and to define common terms, such as what constitutes a “student record.”

Workflows and practices for oversight of learning analytics activities are often deferred to de facto bodies such as institutional review boards. While IRBs are an appropriate entity during research and, in some cases, development, most do not have capacity or obligation to oversee nonresearch applications. In addition, it is often difficult to place learning analytics firmly under the purview of a single institutional governing body. Some institutions have formed learning analytics councils or steering committees to fulfill these roles, though they are typically temporary in nature and therefore lack long-term stability and any ability to provide official oversight. Emerging vice provost–level digital education offices may play a leadership role with respect to learning analytics.[38]

Practical issues for consideration in this domain include whether the institution has a data policy that includes, for instance, the archiving of learning analytics data and back-up and disaster recovery plans; access protocols that dictate who can see, for example, a student's "risk" score and associated information; and monitoring tools for accessibility compliance.

Predictive Learning Analytics Skills

Technical infrastructure, policies, workflows, and practices cannot be efficiently leveraged without institutional capacity. According *to Building Institutional Capacities and Competencies for Systemic Learning Analytics Initiatives*:[39]

> The development of predictive models or adaptive algorithms, particularly those used to make consequential decisions and recommend courses of action to teachers or students, requires contributions from experts with backgrounds in statistics, predictive analytics, learning sciences, measurement, and data visualization. Analytic expertise is needed to develop and validate predictive models; integrate, coordinate, and use data inputs from multiple data systems; help interpret the meaning of significant predictive variables as well as determine appropriate interventions based on the strength of models; and develop data visualization tools and other means of synthesizing data for mass consumption.

Some specific learning analytics skill sets include:

- **Data Science:** This set of skills includes the analysis, interpretation, and visualization of data.
- **Programming:** Programmers to assist with data mining are a fundamental ingredient in moving from models to intervention. This capability may be provided in-house or through a vended product (in this case, however, institutions will still need to work with local integration).
- **Data Literacy:** Data literacy is needed to develop the necessary predictive models and algorithms.
- **Research:** Research expertise is often an imperative to ensure proper understanding and interrelation of highly contextualized and nuanced educational data.
- **Intervention:** Data models and algorithms are the first step, but they alone are not sufficient for ensuring student success in coursework. Intervention planning will identify the right time to intervene, as well as the frequency, method, and tone of intervention that should be adopted.
- **Instructional Design:** Instructional design expertise may need to be leveraged for any number of reasons in predictive analytics. If the goal of a predictive learning analytics model is to determine which part of a course is most difficult for students, the skills of an instructional designer and educational researchers will be invaluable to help close feedback loops, measure outcomes, improve further action steps, and ensure efficiency of resources. Additionally, it may be important to use backwards-design techniques to develop modules or activities to gather more meaningful behavior data to inform predictive models.

Conclusion

Predictive learning analytics, fueled by rich sources of learning data and sophisticated business intelligence tools, clearly holds the potential to revolutionize education. Although still in its early stages, this field will increasingly provide institutions of higher education with new capabilities that can be leveraged by administrators, faculty, staff, and students to address some of the most urgent student-success challenges facing colleges and universities today. Higher education leaders will need to gain knowledge and understanding of the strategic and tactical issues related to the deployment of predictive learning analytics. Such issues span everything from technical challenges in accessing and extracting vital learning data from the many systems on campus and in the cloud to addressing new strategic policy and ethical issues related to the decisions that will be increasingly made based on analytics. Given both the complexity of these issues and the tremendous value and benefits that this technology will bring to our institutions and, most importantly, to our learners, it will become increasingly imperative that we, as a community of IT higher education professionals, continue to work to understand how to best leverage predictive learning analytics and share our collective knowledge across our diverse and global community.

Authors

Special thanks go to the following ECAR-ANALYTICS Working Group authors of this report.

Sakinah Alhadad
Learning Consultant (Research and Evaluation)
Griffith University

Kimberly Arnold (Co-Chair)
Senior Evaluation Consultant
University of Wisconsin–Madison

Josh Baron (Co-Chair)
Assistant Vice President, Information Technology for Digital Education
Marist College

Ilana Bayer
Assistant Professor, Learning Technologies, Program for Faculty Development
McMaster University

Christopher Brooks
Research Assistant Professor, School of Information, and Director of Learning Analytics and Research, Digital Education and Innovation
University of Michigan

Russ R. Little
Chief Innovation Officer
PAR Framework

Rose A. Rocchio
Director, Educational and Collaborative Technologies
UCLA

Shady Shehata
Principal Data Scientist, Business Intelligence and Analytics
D2L Corporation

John Whitmer
Director for Analytics and Research
Blackboard

Additional thanks go to the following ECAR-ANALYTICS Working Group members:

Hendrik Drachsler
Assistant Professor, Welten Institute Research Centre for Learning, Teaching, and Technology
Open Universiteit Nederland

Daniel Huston
Director of Analytics
Rio Salado College

Jocelyn Manderveld
Project Manager
SURF

Robin Ying
Vice President for Information Systems
Tidewater Community College

Citation for This Work

ECAR-ANALYTICS Working Group. *The Predictive Learning Analytics Revolution: Leveraging Learning Data for Student Success*. ECAR working group paper. Louisville, CO: ECAR, October 7, 2015.

Notes

1. According to the 2014 "Education at a Glance: OECD Indicators," "44% of 25–34 year-olds in the United States have a university-level degree (the OECD average is 39%) while the proportion of tertiary-educated 25–34 year-olds is larger in 11 other countries." The United States Country Note that identifies this information also notes, however: "While a large proportion of adults in the United States have university-level education…the tertiary attainment rate is increasing much faster in many other countries." For more on this and what it might mean, see also Goal 2025 from the Lumina Foundation.
2. Although predictive analytics remains at the center of this emerging area of focus, there are intersections with many other disciplines, including the learning sciences and educational psychology, whose theories play critical roles in related issues of curriculum design, instructional methodologies and intervention strategies. We encourage readers to explore them through the additional resources and references that are identified throughout the paper.
3. This distinction originates in George Siemens's early mapping of the domain. For more information, see his August 5, 2011, blog post, "Learning and Academic Analytics."
4. See, for example, Simon Buckingham Shum, "Learning Analytics," UNESCO policy brief, November 2012; "ELI 7 Things You Should Know About First-Generation Learning Analytics," ELI, December 2011; and Malcolm Brown, "Learning Analytics: Moving from Concept to Practice," ELI brief, July 2012. More can be found at the *Wikipedia* page on learning analytics.
5. 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, February 27–March 1, 2011.
6. As in all emerging fields, lessons are still being learned about how predictive learning analytics can be best applied, especially as it relates to student-facing dashboards. Continued study is needed to understand how students understand, reflect on, and internalize dashboards that compare their behavior to that of others.
7. Kimberly Arnold, "Signals: Applying Academic Analytics," *EDUCAUSE Quarterly* 33, no. 1; Kimberly E. Arnold and Matthew D. Pistilli, "Course Signals at Purdue: Using Learning Analytics to Increase Student Success," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, ACM, April 2012, 267–270.
8. See D2L Brightspace Degree Compass.
9. Steven Lonn and Stephanie D. Teasley, "Student Explorer: A Tool for Supporting Academic Advising at Scale," in *Proceedings of the first ACM conference on Learning @ scale conference* (ACM: New York, 2014), 175–176.
10. See Student Success System.
11. Predictive Analytics Reporting: Using Technology to Help Students Succeed, UND blogs, November 18, 2014.
12. See Starfish Retention Solutions.
13. See Eden Dahlstrom and Jacqueline Bichsel, *ECAR Study of Undergraduate Students and Information Technology*, *2014*, research report (Louisville, CO: ECAR, October 2014), available from the 2014 Student and Faculty Technology Research Studies hub. This report notes, however, that only 86% of faculty and 83% of students indicate they use the LMS.
14. Eden Dahlstrom, D. Christopher Brooks, and Jacqueline Bichsel, *The Current Ecosystem of Learning Management Systems in Higher Education: Student, Faculty, and IT Perspectives,* research report (Louisville, CO: ECAR, September 2014), 10.
15. See Family Educational Rights and Privacy Act (FERPA).
16. See ADL Initiative.
17. See IMS Global.
18. Previously called the Tin Can API. For more, see Experience API.
19. See Caliper Analytics.
20. For more information, see the *Wikipedia* pages Semantic Web and Resource Description Framework.
21. See, for example, the Common Education Data Standards from the Department of Education, PAR Framework, and the work of IMS Global. The IMS Learning Tools Interoperability (LTI) specification continues to be the principal method that vendors rely on to exchange learning data.

22. More about "Use of Third-Party (Cloud) Providers of Storage" can be found in Douglas Blair et al., *Research Data Storage: A Framework for Success*, ECAR working group paper, July 15, 2014. In addition, the role of data in the cloud can also be found in the series *Preparing Your IT Organization for the Cloud*.

23. For more on calculating cloud costs see Teri Abbo et al., *TCO for Cloud Services: A Framework*, working group paper, April 24, 2015.

24. Leo L. Pipino, Yang W. Lee, and Richard Y. Wang, "Data Quality Assessment," *Communications of the ACM* 45, no. 4, 2002, 211–218.

25. For more about data security, visit the *Information Security Guide*, which includes content on, for example, data classification, confidential-data handling, anonymization, and much more.

26. Some publicly available instruments may help institutions assess maturity, readiness, and capacity in context. Though these instruments are not specific to predictive learning analytics, they may nonetheless be relevant and of use. The analytics maturity index developed by ECAR measures analytics maturity on six dimensions: process, culture, expertise, investment, governance/infrastructure, and data/reporting/tools The Learning Analytics Readiness Instrument (LARI) is predicated on institutional reflection about readiness to embark upon or grow an analytics initiative. The LARI has five statistically derived domains: governance and infrastructure, ability, data, culture, and process. See Kimberly E. Arnold, Steven Lonn, and Matthew D. Pistilli, "An Exercise in Institutional Reflection: The Learning Analytics Readiness Instrument (LARI)," in *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (New York: ACM, 2014), 163–167.

27. Donald M. Norris and Linda L. Baer, *Building Organizational Capacity for Analytics*, EDUCAUSE, February 2013.

28. Ibid., 31–32.

29. Kyle M.L. Jones, John Thomson, and Kimberly Arnold, "Questions of Data Ownership on Campus," *EDUCAUSE Review*, August 25, 2014.

30. Colleen Marzilli, Julie Delello, Shelly Marmion, Rochell McWhorter, Paul Roberts, and T. Scott Marzilli, "Faculty Attitudes Towards Integrating Technology and Innovation," *International Journal on Integrating Technology in Education* 3, no.1, March 2014, 1.

31. The Open University, "Policy on Ethical Use of Student Data for Learning Analytics," September 2014.

32. Niall Sclater, "Code of Practice for Learning Analytics: A Literature Review of the Ethical and Legal Issues," Jisc, November 2014, 5.

33. See "Code of Practice for Learning Analytics."

34. See Section 508 and Related Laws and Policies.

35. For more information, see the White House report "Big Data: Seizing Opportunities, Preserving Values," May 2014; Dan Solove's blog entry, "Big Data and Our Children's Future: On Reforming FERPA," May 6, 2014; the Department of Education press release "Guidance for Schools Issued on How to Keep Parents Better Informed on the Data They Collect on Students," July 25, 2014; and news regarding an attempt to introduce a new bill to protect student privacy and amend FERPA in Jake Williams, "Senate Bill Attempts to Modify FERPA in Era of Big Data," *FedScoop*, July 31, 2014. In addition, Anya Kamenetz's article "What Parents Need to Know About Big Data and Student Privacy" touches on some of these issues as well as provides some background to these efforts.

36. Douglas Blair et al., *The Compelling Case for Data Governance*, ECAR working group paper, March 17, 2015.

37. More about data stewardship and what is necessary for a data stewardship program is being developed as part of a current ECAR working group project on this topic; the full report is expected in late 2015.

38. See, for example, Digital Education & Innovation at the University of Michigan or MIT's Office of Digital Learning.

39. Kimberly E. Arnold, Grace Lynch, Daniel Huston, Lorna Wong, Linda Jorn, and Christopher W. Olsen, "Building Institutional Capacities and Competencies for Systemic Learning Analytics Initiatives," in *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (ACM, March 2014), 257–260.