In recent years, low background rumblings have been heard in the land of education and training—rumblings that are getting louder each day. These are the sounds of the learning world discovering what Internet professionals working in other market sectors have known for years: The “digital breadcrumbs” that learners leave behind about their viewing, reading, engagement, and assessment behaviors, about their interests, and about their preferences provide massive amounts of data that can be mined to better personalize their learning experiences.1

By Ellen Wagner and Phil Ice

Data Changes Everything: Delivering on the Promise of Learning Analytics in Higher Education
Using evidentiary methods and technology tools to figure out how to interpret what these digital breadcrumbs are telling us has come to be known as learning analytics. Learning analytics provide the tools, technologies, and platforms to empower educators to open the door on meaningful learning experiences that can engage, inspire, and prepare current and future students for success.

The Evolving Learning Analytics Landscape
Technological developments have certainly served as catalysts for the move toward the growth of analytics in business, industry, and education. Data warehouses and “the cloud” make it possible to collect, manage, and maintain massive numbers of records. Sophisticated technology platforms provide the computing power necessary for grinding through calculations and turning the mass of numbers into meaningful patterns. Data mining uses descriptive and inferential statistics (e.g., moving averages, correlations, and regressions) and complex functions (e.g., graph analysis, market basket analysis, and tokenization) to look inside those patterns for actionable information. Predictive techniques (e.g., neural networks and decision trees) help anticipate behavior and events.

The goal of the wide variety of tools for applying descriptive, inferential, and predictive analyses to the massive number of individualized digital records at a transactional level is to improve both general and personalized experiences. This is done to move users/consumers to a desired course of action more quickly. For example, Internet companies study website information to enhance site visitors’ experiences and provide advertisers with more granular targeted advertising. Telecommunications companies mine millions of call-detail records to predict customer churn and profitability. Retailers analyze detailed transactions to better understand customers’ shopping patterns and to forecast demand, optimize merchandising, and increase the lift of their promotions. And although it would be inaccurate to suggest that educational experience is an amalgamation of transactions just waiting to be mined, analyzing student information to better understand learning and motivational patterns may, in fact, offer opportunities for reconsidering how to optimize educational experiences that promote and enable student success.

Pattern Recognition and Business Intelligence Techniques
Unlike many of the research methods employed within the educational research community, business intelligence (BI) methods use analytical techniques to mine data in order to help users make decisions based on real, anticipated, and predicted outcomes. Techniques for analyses include many of the same tools used in academic research settings, but the focus on structuring the methodology is significantly different.

The business intelligence market, although relatively mature, keeps reinventing itself as new technology capabilities enable better performance, better scalability, higher adoption, more pervasive use, and better user experiences. The following are three current areas of innovation for business intelligence technology:

- **Consumerization**, with a technology dependence on search, mobile, visualization, and data discovery
- **Decision support**, emphasizing collaborative decision making and predictive analytics
- **Analysis of nontraditional data and “big data,”** exploring behaviors, transactions, content, text, semantics, and in-memory analytics

These are all emerging examples of an approach called Pattern-Based Strategy (PBS), introduced by Gartner Research in August 2009. PBS guides stakeholders in anticipating new opportunities by helping them recognize the proverbial right place and right time and thus increasing their probability of driving to a point of success. Gartner developed PBS in response to research demonstrating business leaders’ desire to move from reacting to events that had major effects on business strategy and operations to proactively seeking patterns that might indicate an impending event. Interest in the strategy has exploded in the past three years as decision-makers have begun to understand the emerging technologies that can help seek patterns from both traditional (e.g., financial information, inventory, other operational measures) and nontraditional sources of information (e.g., social media, news, blogs). But PBS requires more than an understanding of the technology; it requires cultural changes as well. Finding a pattern is only the tip of the iceberg; Pattern-Based Strategy goes beyond seeking the pattern into modeling the impact of the pattern and adapting the organization to the change required by the pattern. Unfortunately, most organizations do not have a culture of seeking patterns, modeling impact, and rapidly changing the organizational culture to respond appropriately.

Analytics in the Higher Education Enterprise
Pattern recognition and predictive analytics are more commonly used in consumer settings—for example, to help generate recommendations for Amazon and Netflix shoppers or to find prospective dating partners via Match.com. They are not yet broadly used in educational settings, where they could assist with activities such as selecting courses or predicting when students might be at a point of increased academic risk. Educators are still getting their heads wrapped around the realities and implications for what it means to systematically track information à la Google Analytics (http://www.google.com/analytics/) and to use that information in practice. Nevertheless, speculations in the popular and business press about the emerging role and impact of “big data” analyses have ramped up everyone’s expectations of what may be possible for improving accountability,
transparency, and quality. Learning and development organizations simply cannot live outside the enterprise focus on the measurable, tangible results now driving IT, operations, finance, and other mission-critical applications.

As noted in the EDUCAUSE Learning Initiative (ELI) report “7 Things You Should Know about First-Generation Learning Analytics,” learning analytics apply the model of analytics to “the specific goal of improving learning outcomes.” Learning analytics are used to collect and examine the records of students’ interactions with various computer systems and “to look for correlations between those activities and learning outcomes.” The report continues:

The type of data gathered varies by institution and by application, but in general it includes information about the frequency with which students access online materials or the results of assessments from student exercises and activities conducted online. The types of analyses performed vary, but one approach involves the evaluation of historical student data to create predictive models of successful and at-risk students. Reports can take various forms, but most feature data visualizations designed to facilitate quick understanding of which students are likely to succeed.

Lessons from Major League Baseball

As higher education finds itself on the verge of diving deeply into the analytical end of the education transformation pool, Major League Baseball can provide an example of the impact that data can have on informing decisions. The recent movie Moneyball, based on the 2003 book Moneyball: The Art of Winning an Unfair Game, tells the story of how the Oakland Athletics applied the principles of what became known as “sabermetrics” to analyze every aspect of their game and to then invest their relatively scant salary dollars in very smart, statistically significant ways. It is not a story of how statistics saved the day for Oakland, any more than collecting more and more data on everything we do in higher education is going to save the day for colleges and universities. Statistics have always been collected on just about anything that anybody did, is doing, and might do in baseball; one might suggest that thanks to today’s student information systems and learning management systems, the same can be said of U.S. postsecondary educational enterprises. What was different in this case was that the Oakland A’s started using sabermetrics to analyze each and every aspect of the game—and the business—of baseball. The A’s executive leadership then made decisions informed by the specialized analysis of objective, empirical evidence, specifically baseball statistics measuring in-game activity.

Bill James, one of the pioneers in sabermetrics and a prominent advocate, was puzzled that in spite of evidence to the contrary, important baseball decisions continued to be made based on personal biases and folk wisdom. He was also struck by the aversion to using the data that the baseball industry collected about itself: “Baseball keeps copious records, and people talk about them and argue about them and think about them a great deal. Why doesn’t anybody use them? Why doesn’t anybody say, in the face of this contention or that one, ‘Prove it?’”

Of course in education, the methodological need to focus on the tenability of a particular research finding at a particular level of significance will trump naïve calls to “prove” anything. Nevertheless, the strong desire for proof burns bright in education, particularly
Learning from the PAR Framework

Unlike the descriptive data that emerges from compilation and analysis tools like those used in Google Analytics, the Predictive Analytics Reporting (PAR) Framework offers another approach to learning analytics. Developed by the WICHE (Western Interstate Commission for Higher Education) Cooperative for Educational Technologies (WCET), the PAR Framework was modeled along “big data” approaches to predict points of student loss and momentum. It is one example of how some educational researchers are starting to look “beyond the null hypothesis” for retention and completion guidance. The desire to find a new way to deconstruct risks to student success comes from a position of great need. Despite high enrollment numbers, postsecondary completion rates have generally remained unchanged for the past forty years. Furthermore, of all students who enroll in postsecondary education, less than half (46.1%) attain a degree within 150 percent of “normal time” to degree. Although online learning has offered a legitimate path for pursuing a college education and provides students with a convenient alternative to face-to-face instruction, it is laden with retention-related concerns.

The PAR Framework was conceptualized as a way to take advantage of current and emerging decision-making approaches that use business intelligence techniques to look for patterns that would not be evident unless viewed through a single, multi-institutional database. The PAR Framework team first tested pragmatic, evidence-based frameworks that had been implemented and refined, independently, at the American Public University System, Rio Salado College, and the University of Phoenix. These institutions shared their knowledge and expertise to extend the existing frameworks into a multi-institutional approach. Phil Ice, principal investigator of the PAR Framework Project, has posited: “Unifying records from multiple schools has the potential to expose generalizable patterns that can provide guidance to institutions across the board, regardless of their internal level of analytics expertise.”

Eventually, six major institutions including for-profits, research universities, and community colleges aggregated their data, representing more than 640,000 student records, with over 3 million course-level transactions.

Six major institutions including for-profits, research universities, and community colleges aggregated their data, representing more than 640,000 student records, with over 3 million course-level transactions.

Six major institutions including for-profits, research universities, and community colleges aggregated their data, representing more than 640,000 student records, with over 3 million course-level transactions.

With calls for accountability and evidence dominating current discussions.

Six major institutions including for-profits, research universities, and community colleges aggregated their data, representing more than 640,000 student records, with over 3 million course-level transactions.

Thirty-three common variables were identified and collected from the participating institutions and were normalized across the institutions, resulting in a single consistent framework with all variables being commonly defined. Student activity/inactivity was assessed at six to eight months after the collection of data about student course activity. Students who were no longer enrolled and had not graduated were defined as disenrolled and at-risk for degree completion. Further analysis revealed several promising results:

- identify common variables influencing student retention and progression;
- establish factors closely associated with online students’ proclivity to remain actively enrolled within the institution;
- determine if measures and definitions of retention, progression, and completion differ materially among various types of postsecondary institutions; and
- discover advantages and/or disadvantages to particular statistical and methodological approaches pertaining to identifying profiles of students considered to be “at-risk.”

For students at-risk, disenrollment was influenced by the number of concurrent courses in which students were enrolled: taking more than one course in the early stages of their college career was highly correlated with an increased risk of disenrollment.
No apparent relationships existed between age, gender, or ethnicity as a function of the student's risk profile.

For students not at-risk of disenrollment, institution-specific factors predicted student success.

These initial findings suggest there are many opportunities for better identifying and serving prospective students by focusing on the intersection of student needs and institutional uniqueness in order to find students' individual best-fit with an institution. Overall, results from the PAR Framework proof-of-concept project offer compelling evidence that analyses of the normalized variables tracked in a multi-institutional database of student records can provide meaningful benchmarks for exploring loss and momentum at the student and course levels.

What We Know for Sure about Learning Analytics

The 2012 Horizon Report, produced by the New Media Consortium and the EDUCAUSE Learning Initiative, reflected on the impact of “Learning Analytics” as an emerging technology, projecting its time-to-adoption as two to three years. However, we already know several things for sure.

Analytics are here today, and they are here to stay. There is no question that information can be gleaned from the transactions and interactions we leave behind along the paths of our online lives. This information can then be summarized in reports and displays that provide intelligence for making better-informed decisions to shift patterns of behaviors in desirable ways. Analytics are already being used in a variety of ways in the higher education enterprise. Operations, finances, student recruitment, and planned giving represent arenas where business intelligence techniques are starting to be used with greater frequency. Interest in learning analytics continues to explode as more and more tools emerge, more techniques are validated, and our collective understanding of analytics utility continues to grow. Even so, we are still on the early side of the analytics adoption curve, especially when compared with other U.S. economic market segments such as retail, telecommunications, financial services, and manufacturing.

Analytics are a means to the end, not the end in and of themselves. The point of analytics is to enable better decision making. People still need to make the decisions.

Research on analytics impact and efficacy is essential. The more often we hear the calls for applying learning analytics techniques at various points in the teaching and learning process, the greater is the responsibility of researchers to ensure that the techniques being promoted and the reports they engender do, in fact, produce valid, reliable, and repeatable results. The leading-edge work of the Society for Learning Analytics Research (SoLAR) and other like-minded research communities is essential to an understanding of what works and what does not.

Conducting research to explore various dimensions of learning analytics is a fundamentally different undertaking from analyzing learning measures and outcomes to look for patterns that can inform decision making about improving student success.
We need to give front-line educators the knowledge, skills, and tools to use data to inform decision making. Data, by itself, does not improve student success. Although learning analytics offer great promise for transforming the accountability, personalization, and relevance of the 21st-century postsecondary educational experience, that promise will not be fully realized until we put the power of better-informed decision making into the hands of front-line educators.

Common data definitions will be required if we intend to compare “apples to apples” in assessing points of student loss and momentum. Evidence from the PAR Framework suggests that a shared data model and common definitions that have been negotiated and agreed upon between and across institutions are needed to conduct “big data” styled analyses that will foster multi-institutional utility in higher education. Beth Davis, director of the PAR Framework, has noted that the PAR Framework’s common data definitions afforded a unique opportunity for diagnosing conditions across, between, and within institutions: “What we saw in the PAR Framework analyses work was that, without a common diagnosis, it will be difficult to agree on treatment to cure the diagnosed problem. Without agreed-upon treatments used in response to a common diagnosis, it is difficult to measure efficacy of treatments. If one cannot measure efficacy, it is impossible to scale. And perhaps worst of all, we continue to guess about what really works in the service of student success.”

There’s no such thing as “sort of” transparent. Once one can see the results of the analyses being run against any number of institutional data sources, including student achievement data, it will be increasingly difficult to ignore what the numbers are saying.

We need to be sure educators are ready to live under the “sword of data.” When Douglas Bowman, the lead designer at Google, discovered that his design decisions were being overruled by engineers fueled by customer-preference statistics, he resigned in protest against working under what he called the “sword of data.” He declared that he had “grown tired of debating such minuscule design decisions” after a team at Google couldn’t choose between two blues. They decided to test the two colors against forty-one shades to see which color performed better. “I had a recent debate over whether a border should be 3, 4 or 5 pixels wide, and was asked to prove my case. I can’t operate in an environment like that... I won’t miss a design philosophy that lives or dies strictly by the sword of data.” How will educators respond to growing expectations around data-driven decision making when their “art of teaching” may be confounded by empirical evidence to the contrary? Is education on the verge of experiencing its “Moneyball moment”? Will learning analytics lead us past the brink of “Wal-martification”?

We haven’t even begun to scratch the surface of the possibilities. Using data to diagnose problems is only part of the opportunity. Analytics will contribute to informed decision making to more and more parts of the educational enterprise. Several recent examples from education provide evidence that the community is starting to pay attention to the opportunities for leveraging learning analytics in the service of high-quality, scalable, personalized learning experiences. For example, Craig Powell, CEO and founder of ConnectEDU, has suggested that an online Match.com-like college admissions service could transform the admissions process, while Austin Peay State University students have begun using a recommendation system, designed by
their provost, to get help picking their courses—a step that could change GPAs and career paths.12

Conclusion
As the postsecondary education community grapples with changing social conditions, with demands from new and growing stakeholder groups, and with ongoing assaults to familiar sources of funding, the perceived power of data-driven decision making continues to alarm, provoke, seduce, and intrigue. Emerging evidence from research and practice communities suggests that learning analytics may enable learning experiences that are more personal, more convenient, and more engaging and may also have a direct positive impact on student retention. Learning analytics have the potential to help learners and instructors alike recognize danger signs before threats to learning success materialize.

These possibilities provide the fuel that is feeding the fires of increasing interest in learning analytics. Although effective practices leveraging learning analytics are still in their early stages, the number of stakeholder groups interested in discovering more about how learning analytics can support effective decision making in higher education institutions continues to grow. By using learning analytics and working together to make accountable, informed decisions based on carefully mined data, all contributors to the educational enterprise—students, staff, faculty, administrators, and members of the broader higher education community—can help deliver on the promise to enable student success.

Notes
9. Beth Davis, personal communication with the authors, May 18, 2012.

© 2012 Ellen Wagner and Phil Ice. The text of this article is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Ellen Wagner (edwsonoma@gmail.com) is Executive Director of the WICHE (Western Interstate Commission for Higher Education) Cooperative for Educational Technologies (WCET) and is Partner and Senior Analyst with Sage Road Solutions, LLC.

Phil Ice (pice@apus.edu) is Vice President of Research and Development at American Public University System.