



Scaling Student Success with Predictive Analytics: Reflections After Four Years in the Data Trenches

Ellen Wagner & David Longanecker

To cite this article: Ellen Wagner & David Longanecker (2016) Scaling Student Success with Predictive Analytics: Reflections After Four Years in the Data Trenches, *Change: The Magazine of Higher Learning*, 48:1, 52-59, DOI: [10.1080/00091383.2016.1121087](https://doi.org/10.1080/00091383.2016.1121087)

To link to this article: <http://dx.doi.org/10.1080/00091383.2016.1121087>



Published online: 07 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 313



View related articles [↗](#)



View Crossmark data [↗](#)

SCALING STUDENT SUCCESS WITH PREDICTIVE ANALYTICS

Reflections After Four Years in the Data Trenches

By Ellen Wagner and David Longanecker

Ellen Wagner (ellen.wagner@parframework.org) is chief research officer and co-founder of the PAR Framework, a non-profit data cooperative focused on institutional effectiveness and student success. She was vice president of WICHE and led that organization's Cooperative for Educational Technology from 2009 through 2013. David Longanecker (dlonganecker@wiche.edu) is the president of the Western Interstate Commission for Higher Education (WICHE).

In Short

- The metrics used in the US to track students do not include adults and part-time students. This has led to the development of a massive data initiative that uses predictive analytics to trace the progress of all types of students in the system.
- The Predictive Analytics Reporting (PAR) framework provides actionable, institution-specific insight about what affects undergraduate student success and failure to member institutions from two- and four-year public, proprietary, traditional, and progressive institutions. The institutions then exchange information about the tools and practices that they have found measurably improve student outcomes. Collaboration and leadership are vital to the identification and use of those tools and practices.
- Viewing normalized data through a multi-institutional lens and using a common set of measures with common data definitions leads to insights that are not available when looking at records from a single institution.
- PAR helps members eliminate redundant programs, understand the scale of their programs, match interventions with the causes of student academic risk, measure the impact of student-success programs, and respond to budget cuts with informed decisions.
- PAR also helps member institutions remedy their local data problems; organize data across silos; uncover gaps, errors, and overlaps in student-data elements; and compare themselves to other campuses.

Metrics currently used to describe and compare the performance of colleges and universities in the United States do not include the post-traditional students, instructional methods, business models, and data resources that distinguish contemporary higher education. This deficiency has led to the development and launch of a massive data initiative that uses predictive analytics to track patterns of momentum and loss for all types of students in the US postsecondary system.

This initiative, named the Predictive Analytics Reporting Framework (PAR), is providing insight into the performance of a wide variety of student types, including part-time and adult students, all of whom are more likely than full-time, residential students to pursue degrees and certificates through for-profit and alternative delivery models such as online and competency-based learning programs. But PAR is a data resource for analyzing traditional as well as non-traditional student populations.

Predictive analytics are methodologies, data, statistical algorithms, and machine-learning techniques to identify the likelihood of future outcomes based on historical data (http://www.sas.com/en_us/insights/analytics/predictive-analytics.html). They have been used for business intelligence and data mining in a wide variety of commercial and governmental settings over the past decade. They are increasingly being used in educational settings to anticipate future opportunities or to remove barriers to success by finding students at risk of dropping out of college.

By showing non-traditional and traditional students together in the same data set, with a focus on outcomes achievement, PAR has also started to identify potential improvements to federal data collection and statutory disclosure and reporting requirements. Current student progress scorecards often focus on the traditional college student: they are based upon the records of students who are attending college for the first time, are going to school full-time, and are living on campus.

PAR focuses on outcomes data for *all* students, including those who are non-residential, part-time, in the military, or completing progressive programs at a distance or online. This reflects the legitimacy of all segments of the higher educational student population as participants in the pursuit of a college education. But perhaps of greatest potential value is PAR's current work to determine the effectiveness of interventions that aim to keep students in college.

PAR'S HISTORY

PAR began as a big audacious idea back in 2010, when members of the Western Interstate Commission for Higher Education's (WICHE's) Cooperative for Educational Technology (WCET) began speculating about the value of using predictive analytics to address the ongoing problem

of student attrition in US postsecondary education. Despite much investment in strategies designed to increase student success, postsecondary education completion rates have remained virtually unchanged for the past 40 years. Of all students who enroll in postsecondary education, less than half (46.1 percent) complete a degree within 150 percent of "normal time" to degree (Knapp, Kelly-Reid, & Ginder, 2012).

At the time, WCET members were particularly interested in what could be done to improve the retention and completion rates of online students. While online learning offers a legitimate path for pursuing a college education and provides students with a convenient alternative to face-to-face instruction, it has even lower retention and completion rates than traditional programs.

PAR began with six forward-thinking postsecondary institutional partners contributing anonymized and de-identified student and course data into a dataset managed by WCET, which built predictive models, managed the data, and directed all project operations (Ice et al., 2012). These early collaborators determined the factors contributing to the retention, progression, and completion of online learners. Their specific purposes were to reach consensus on a common set of variables for constructing the predictive models and conducting analyses, as well as to explore the advantages and disadvantages of various statistical and methodological approaches to assessing factors that contribute to student retention, progression, and completion.

PAR FRAMEWORK FOUNDING AND IMPLEMENTATION PARTNERS 2011-2014

American Public University System*
Ashford University
Broward College
Capella University
Colorado Community Colleges Online*
Excelsior College
Kaplan University
Lone Star Community College System
Northern Arizona University
Penn State World Campus
Rio Salado College*
Sinclair College
Troy University
University of Central Florida
University of Hawaii System*
University of Illinois-Springfield*
University of Maryland University College
University of North Dakota
University of Phoenix*
Western Governors University

* = Founding Partner

With the results of this initial study as evidence, in the two years that followed the PAR team refined its predictive modeling and descriptive benchmarking efforts, adding 14 colleges and universities to the collaborative and inserting 44 more variables into the dataset.

Benchmarking is the process of comparing one's performance metrics to those of the best-performing other enterprises in order to identify the practices that have generated those results. In postsecondary education, benchmarking compares an institution with others sharing similar characteristics on dimensions of performance and efficiency or on overarching outcomes such as student progress and completion.

Then PAR developed 1) institutional models for finding students at risk; 2) national benchmarks showing comparative outcomes data; and 3) a platform for inventorying, tracking, measuring, and managing interventions to enhance the likelihood of student success. After receiving three research grants from the Bill & Melinda Gates Foundation between 2011 and 2014, PAR conducted rigorous development, testing, and evaluation of its predictive models, benchmarks, and intervention return-on-investment (ROI) tools.

In 2014, PAR added 18 self-funded institutions to the collaborative to explore the feasibility of measuring the

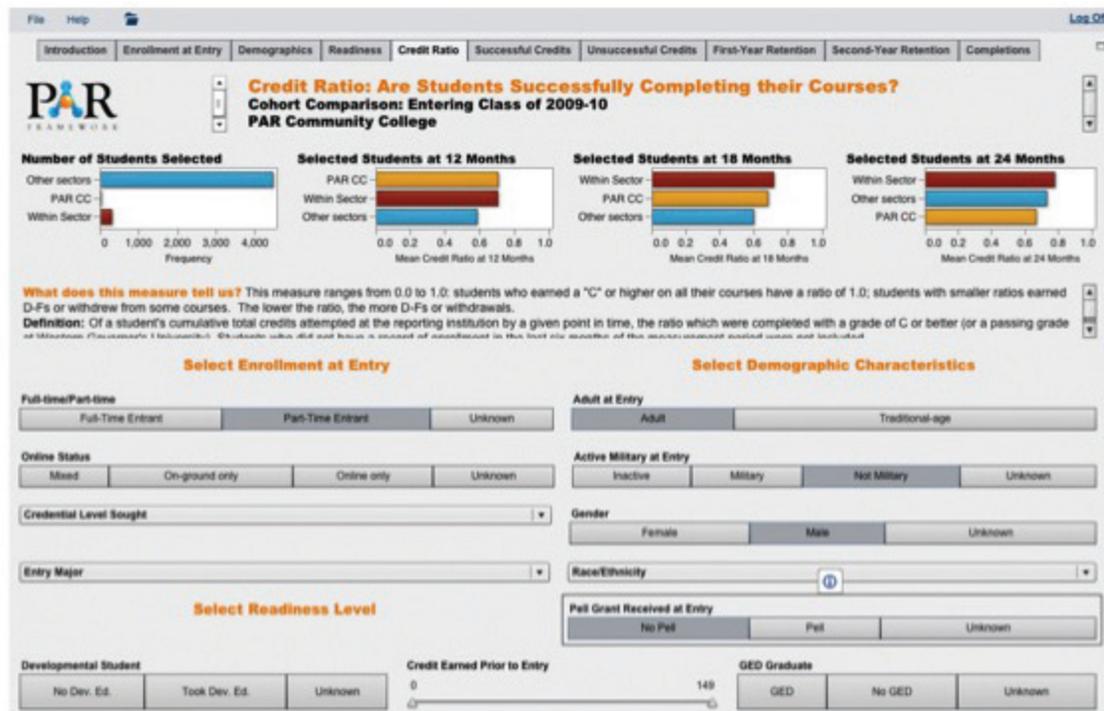
effectiveness and impact of IPAS (integrated planning and advising) platforms. IPAS is an acronym used to describe the category of technology and applications related to academic planning and advising. This includes but is not limited to watchlists, academic planners, program planning and pathway applications, and career-counseling platforms.

Through a collaborative research effort with the American Institutes for Research (AIR), in 2015 another 12 institutions joined PAR to consider extending current NCES academic program attributes and to evaluate how for-profit institutions differed on program and student-specific data points.

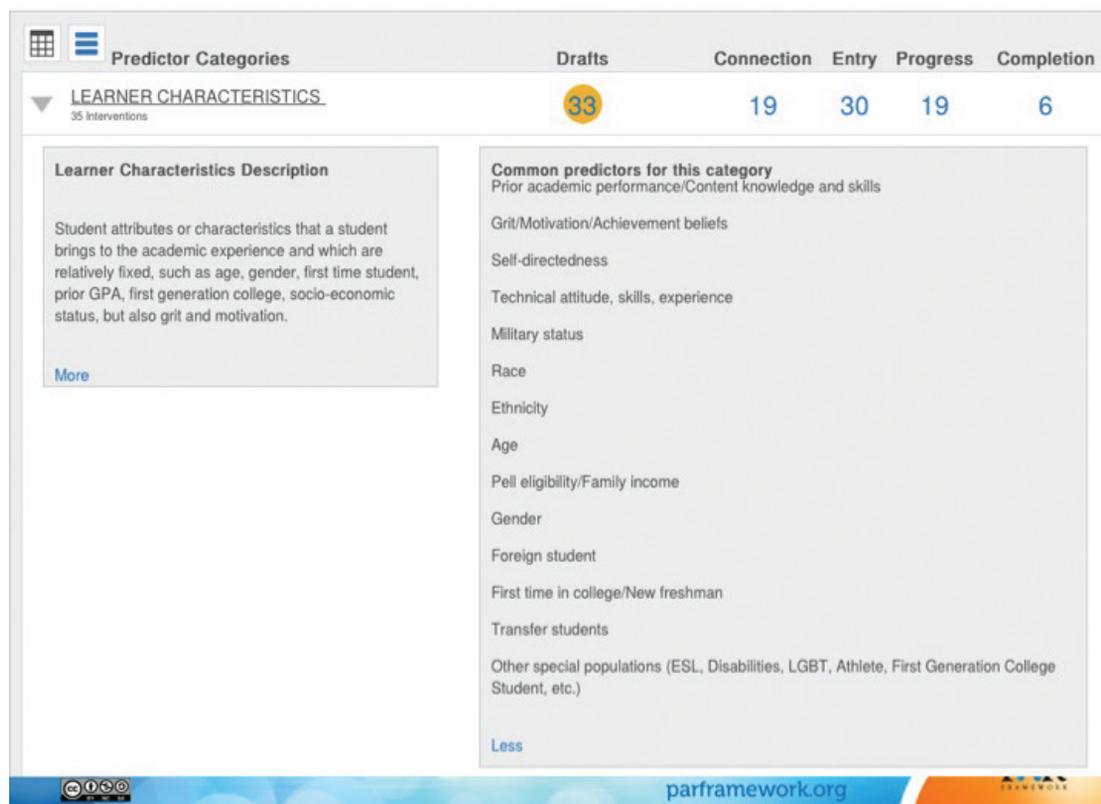
PAR separated from WICHE in January 2015 and became a self-standing, non-profit provider of analytics services. PAR is now included among the Institute for Higher Education Policy (IHEP)'s PostsecData Collaborative National Voluntary Data Projects. Gartner Research (2014) observes that PAR is distinguished from the many data-analytics models emerging in the education domain by its common, openly published data definitions and student-success frameworks.

PAR further differentiates itself by connecting predictions of risk to interventions that have been shown to work with specific risks associated with specific students at specific points in the college-completion life cycle, as measured by improved retention. As Bloemer et al (2014) note, the identification of students at risk is of greater value when tied to interventions that have been empirically shown to mitigate risks for "students like them" at specific point of need.

PAR OUTCOMES BENCHMARKS



PREDICTORS: LEARNER CHARACTERISTICS



How PAR OPERATES

PAR currently works with approximately 2 million anonymized student records and 20 million institutionally de-identified course-level records from more than 350 campuses among its members. It provides actionable, institution-specific insight about the factors that affect undergraduate student success and failure to member institutions from two- and four-year public, proprietary, traditional, and progressive institutions. Participating institutions then exchange information about the tools and best practices that they have found measurably improve student outcomes.

While attention was initially paid only to online students, it became clear in the first few weeks that any study focused on student success would need to deal with the needs of *all* students, not just those who took advantage of technology-based course delivery. The sample now includes the records of all students, including those from on-the-ground, blended, and online programs who attend member institutions.

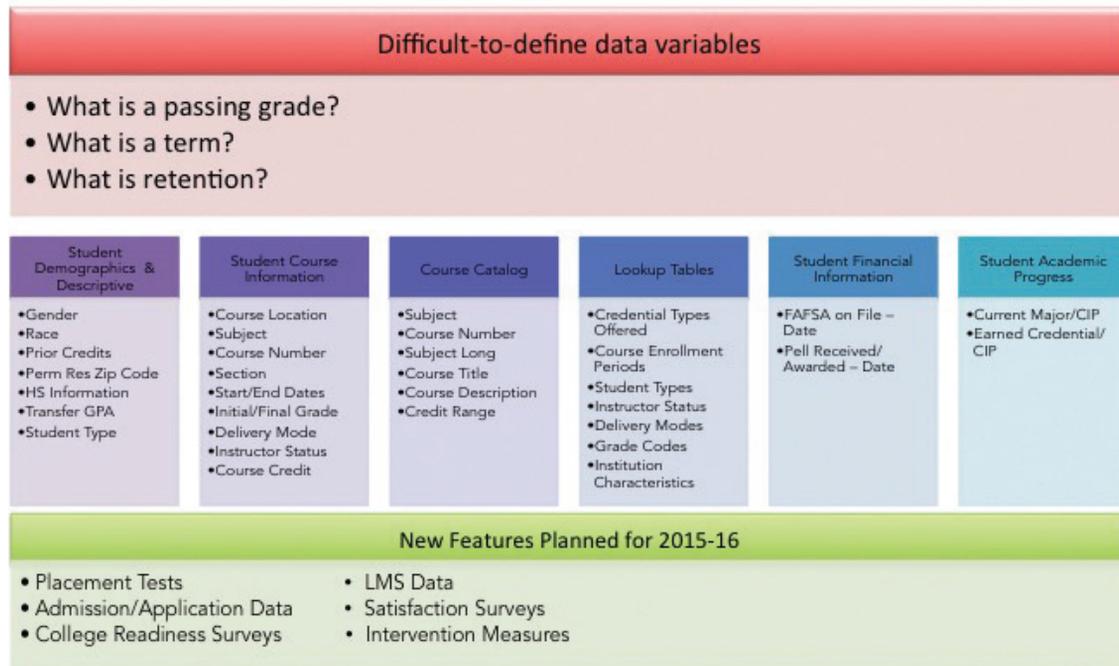
PAR focuses on 77 student variables that are available for each student in the massive data set. Viewing normalized data through a multi-institutional lens and using complete sets of undergraduate data based on a common set of measures with common data definitions (find them at <https://community.datacookbook.com/public/institutions/par>) leads to insights that are not typically available when looking at records from a single institution.

PAR's institutional members gather data according to the PAR Framework's common data definitions and detailed file specifications. Before submitting data, institutions remove any personally identifiable information and replace it with a PAR student ID. Institutions maintain a translation table of their internal IDs to PAR Student IDs, which is used to re-identify those students after the data have been analyzed by PAR.

PAR puts the data through more than 600 quality-assurance tests before including them in its Amazon Web Services-hosted data warehouse. The data are then analyzed to develop institution-, program-, course- and student-level descriptive analytics and predictive insights displayed in predictive-analytic dashboards and in national-benchmark reports built using SAS Visual Analytics (a choice made thanks to unlimited institutional visualization software licenses at PAR member institutions). The members provide incremental data updates at the end of each term/course-enrollment period to measure changes over time, evaluate the impact of student-success interventions, and enable the predictive models to be fine-tuned.

PAR's data experts work with member institutions, providing individualized support for gathering, understanding, and delivering longitudinal student-level data from across institutional systems, using a well-validated set of data definitions and file specifications. The processes and support—combined with scalable, automated quality-assurance tools—

PAR COMMON DATA ELEMENTS



guide member institutions in crafting and delivering accurate and meaningful student-level record sets. Throughout the process PAR data managers help college and university staff diagnose and correct cumbersome and potentially costly institutional data issues that can impede correct reporting, insight, and access to funds based on performance funding and student financial and veteran aid.

PAR's framework for gathering student-level data based on common definitions helps member institutions isolate and remedy their local data problems, such as anomalies in student-cohort reporting. It helps them organize data across systems and silos in order to create longitudinal student-level record sets using common definitions and data types. For instance, PAR helps campuses improve the capture and reporting of student military and veteran statuses across the multiple systems where those data are recorded. In this way, institutions can uncover gaps, errors, and overlaps in student-data elements and make campus-level comparisons possible.

LINKING PREDICTIONS TO ACTION

PAR's openly published data framework and the PAR Student Success Matrix (SSMx) application use a validated mechanism to inventory student-success activities across the institution (see below for a sample of typical predictors). In working with member institutions, PAR staff has found between 80–100 student-success services in effect at any one time on any given campus.

The PAR SSMx helps institutional members

- eliminate redundant programs. Most campuses find that from 10 to 30 percent of their intervention programs are serving the same audience and goal.

- understand the scale of their student-success programs. Many such initiatives are upside-down in terms of the institutional resources attached to the program relative to the number of students served. The SSMx helps institutions right-size their investments to the student need and potential impact on retention and graduation.
- match interventions with the causes of student academic risk. Together with PAR predictive models (customized for each institution) that identify which students are at risk and why, the SSMx specifies which risk factors lack any interventions. For example, while low GPA and course withdrawals contribute to student attrition, many campuses lack initiatives that flag and address those behaviors.
- measure the impact of student-success programs. Even at the most data-driven institutions, only about 10 percent of the intervention programs are evaluated for effectiveness—millions are invested campus-wide with only a limited understanding of their ROIs. The SSMx enables institutions to measure the investment and number of students reached for every intervention. More importantly, PAR analysis measures intervention effectiveness. Readers are encouraged to navigate to the middle of the web page at <http://www.parframework.org/video-testimonials>, where they can view videos of PAR members reflecting on their PAR experiences.
- respond to budget cuts with informed decisions about the fat vs. bone. With a comprehensive understanding of programs and their impact, institutions can make informed decisions on how to eliminate waste and redundancy during times of budget contraction with an assurance that they are cutting the right programs.

INTERVENTION COVERAGE AND GAPS

Military status				
Intervention	Connection	Entry	Progress	Completion
Non-Traditional Knights Website		✓	✓	
Veterans Academic Resource Center	✓	✓	✓	✓

Ethnicity				
Intervention	Connection	Entry	Progress	Completion
McNair Scholars Program			✓	✓

Age				
Intervention	Connection	Entry	Progress	Completion
Non-Traditional Knights Website		✓	✓	

Gender				
Intervention	Connection	Entry	Progress	Completion
Brother to Brother Program		✓	✓	
Girls EXCELing in Math and Science (GEMS)		✓	✓	

REFLECTIONS AFTER FOUR YEARS IN THE DATA TRENCHES

Collaboration

All of the institutions involved with PAR were acknowledged innovators, with at least 10 of the founding 16 members already involved in exploratory to advanced predictive-analytics efforts on their own. But each also recognized the power of collaboration at a time when postsecondary education is being held to higher levels of accountability by many more stakeholders than in the past.

Founding partners also saw safety in numbers while trying to change the conversation about the value and impact of programs serving online, part-time, transfer, and non-traditional students at for- and not-for-profit institutions. Predictive models that were localized to each institution but normed to the aggregated dataset gave institutions a chance to compare and contrast their programs in ways they had never been able to do before and thereby enabled benchmarking. Having a shared mission of improving student success, regardless of the kind of institution or the source of its funding, has continued to serve the collaborative.

The support of senior leadership increased the chances of PAR participation and assured participants that the time spent on doing this work was viewed positively. Having the support of a chief information officer enabled them to complete essential operations such as securing access to data sets and ensuring that timely data pulls remained a priority.

The founding partners were essential for validating the methods and models used for building scalable, repeatable tools and models. Doing the work in public built trust and

encouraged participants to share problems, ideas, and challenges, thereby contributing to the larger national conversation about what data can tell us about student success and how to promote it.

Research

Focusing on research questions kept our efforts targeted on the declared goal of improving student success. By starting this work as a research project, we were not driven by the kinds of demands faced by commercial entities that need to respond to shareholder or investor requirements.

We focused instead on issues of academic progress, retention, and completion that we could generalize to all students regardless of their age; their enrollment status; whether they were taking blended, online, or on-the-ground courses; or whether they attended a for-profit or a not-for-profit institution. This was possible because of the common definitions, the focus on generalizability and replicability, and the ability to openly share our work.

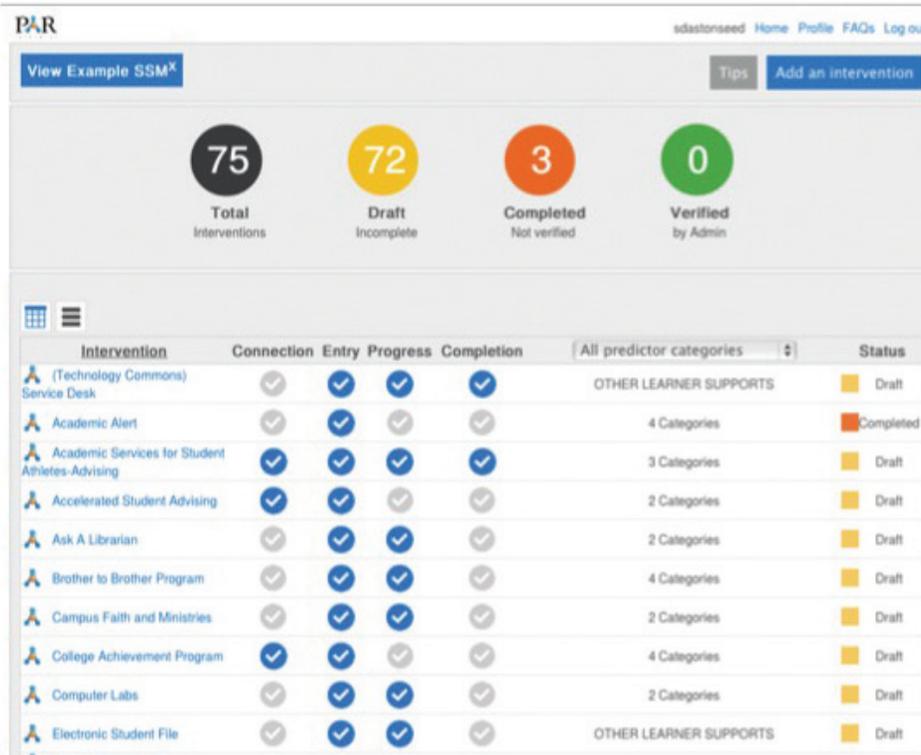
PAR STUDENT SUCCESS MATRIX

Prediction and Intervention

Predicting who and why someone is at academic risk is necessary but insufficient for responding to the complexities of postsecondary student success. Predictions that are not linked to treatments that have been shown to make a difference for the diagnosed risk are empty exercises.

The development of the SSMx was a direct response to questions we received from our institutional members when we shared the results of the predictive analyses telling them

PAR STUDENT SUCCESS MATRIX



which of their students were likely to be unsuccessful in various courses and programs. Before we undertook this work, few institutions inventoried the interventions they had used to improve student performance, persistence, retention, and completion at various points in the course-completion life cycle.

While data professionals will be needed to help construct new modeling and analysis techniques for the future, *everyone on campus*—advisors, faculty, administrative staff, and students—needs to increase their fluency with data-driven decision-making if we are to achieve institutional improvement and student success.

RESOURCES

- Bloemer, B., Swan, K., Cook, V., Wagner, E.D., & Davis, B. (2014). *The Predictive Analytics Reporting Framework: Mitigating academic risk through predictive modeling, benchmarking, and intervention tracking*. Bloomington IL: Illinois Education Research Conference.
- Ice, P., Diaz, S., Swan, K., Burgess, M., Sherrill, J., Huston, D., & Okimoto, H. (2012). The PAR Framework proof of concept: Initial findings from a multi-institutional analysis of federated postsecondary data. *Journal of Asynchronous Learning Networks*, 16(3).
- IHEP. (2015). POSTSECDATA: Mapping the data landscape: An interactive website at <http://www.ihep.org/research/initiatives/postsecondary-data-collaborative-postsecdata>
- Knapp, L.G., Kelly-Reid, J.E., & Ginder, S.A. (2012, December). *Enrollment in postsecondary institutions, fall 2010; financial statistics, fiscal year 2010; and graduation rates, selected cohorts, 2002–2007*. Washington, DC: US Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). Retrieve at <http://nces.ed.gov/pubs2012/2012174rev.pdf>
- Lowendahl, J.M. (2014, July 23). *The education hype cycle, 2014*. Stamford, CT: Gartner Research. Retrieve at <https://www.gartner.com/doc/2806424/hype-cycle-education>
- Wagner, E.D., & Davis B.B. (2013, Dec. 6). *The Predictive Analytics Reporting (PAR) Framework*. Boulder, CO: WCET. Retrieve at <http://www.educause.edu/ero/article/predictive-analytics-reporting-par-framework-wcet>

Become More. Become Extraordinary. At the University of South Dakota.



Downloaded by [4.14.235.30] at 11:26 17 August 2017

Earn your degree in **Educational Administration** at a remarkably affordable price.

Our School of Education is fully accredited by NCATE — the highest professional standard in education.

Choose from unique course offerings with emphasis areas in:

- Student Affairs
- Adult Education/Non-school Settings
- Training and Development

Apply today: www.usd.edu/onlinema-edad



UNIVERSITY OF
SOUTH DAKOTA

**DIVISION OF CONTINUING
AND
DISTANCE EDUCATION**

414 East Clark Street
Vermillion, SD 57069
800-233-7937 | 605-658-6140
cde@usd.edu