NASPA Student Affairs Administrators in Higher Education

April 2017

PREDICTIVE ANALYSIS OF STUDENT DATA

A Focus on Engagement and Behavior

Michelle Burke | Amelia Parnell | Alexis Wesaw | Kevin Kruger

NASPA conducted this research with funding from Lumina Foundation. The views expressed in this publication are those of the authors and do not necessarily represent those of Lumina Foundation, its officers, or employees.

EXECUTIVE SUMMARY

With public scrutiny over the value of higher education increasing, colleges and universities are turning to business intelligence practices to improve outcomes. In higher education, institutional performance is often centered on student enrollment and retention, with an ultimate goal of students' timely persistence to a college degree. As a result, colleges and universities are considering how to use data to intervene proactively with students who are at risk for poor academic performance or low institutional engagement. Many institutions have adopted data analytics practices to forecast operational needs and enrollment trends, and are now applying the use of predictive analytics directly to student success initiatives.

> **"Many institutions have adopted data analytics practices to forecast operational needs and enrollment trends, and are now applying the use of predictive analytics directly to student success initiatives."**

Prior research on predictive analytics in higher education examined the prevalent uses of data and the level of support for overall institutional analytics as well as "learning analytics" related to student success (Dahlstrom, 2016; Yanosky & Arroway, 2015). Although these studies addressed institutional applications of analytics, there was limited detail regarding the factors that influenced institutional support for the use of predictive analytics to increase student retention and persistence. It also appears that a gap in the literature exists regarding how student engagement data are being used in predictive analytics initiatives. Available research suggests that early efforts related to data analytics have focused primarily on academic and learning management system variables (Arroway, Morgan, O'Keefe, & Yanosky, 2016; Dahlstrom, 2016; Yanosky & Arroway, 2015), and while student affairs professionals are called on to implement intervention strategies after at-risk students are identified, student engagement data are not often included in predictive models.

NASPA—Student Affairs Administrators in Higher Education conducted a landscape analysis of the use of predictive analytics by student affairs professionals at higher education institutions. Most student affairs divisions are collecting student engagement data and conducting needs, process, and outcomes assessments. NASPA's research addressed the kinds of student engagement and behavioral data that are collected within student affairs departments and the extent to which institutions are using such data in predictive analytics models. The research also addressed the factors that influence institutions' development of data analytics projects and how various resources are employed to collect the data and conduct the analyses.

"NASPA's research addressed the kinds of student engagement and behavioral data that are collected within student affairs departments and the extent to which institutions are using such data in predictive analytics models."

NASPA interviewed professionals from student affairs, academic affairs, institutional research, information technology, and assessment at 25 colleges and universities to discuss the range of methods for utilizing data analytics to inform retention efforts. The institutions differed in size and approach; however, the interviews revealed that all are committed to student retention efforts and are using or plan to use some form of predictive analytics. Five of the institutions are in the planning phases of predictive analytics projects, 9 have implemented projects within the past 4 years, and 11 institutions have been using some form of data analytics related to student success for 5 years or more. Conversations with members of these institutions yielded several common factors regarding the use of predictive analytics, many of which pertain to the alignment and allocation of personnel and financial resources, including the following:

- institutional commitment to increasing undergraduate retention and improving enrollment management;
- senior-level leadership encourages data-informed decision making;
- strong partnership between campus functions, particularly information technology and institutional research;
- adequate allocation of resources for staff to effectively address the findings produced from predictive models;
- continuous training and support for personnel who collect, analyze, or utilize data;
- capacity to connect data across systems or within one system; and
- increased accountability metrics, such as performance-based funding.

Most of the institutions in the study are still primarily focused on using academic data in predictive models. However, the range of student engagement data that could be used is much broader and could lead to deeper understanding of keys to student persistence. Although most of the institutions could do more to incorporate engagement and behavioral data into their predictive models, they are using these data in the execution of early alert systems, which are retention tactics that target at-risk students for intervention through a variety of support systems. Early alert systems utilize several types of data, including pre-enrollment variables such as high school grade point average and standardized test scores, academic variables such as mid-term grades and course attendance, motivation and self-efficacy variables such as students' self-reports of connectedness to the institution, use of support services such as advising and tutoring, and student engagement variables such as participation in campus activities.

One challenge for many institutions is limited capacity to gather accurate student engagement and behavioral data and connect them to the student information system for inclusion in predictive models. Several institutions are strategically planning how to meet that challenge with improved data collection and less siloed data analysis. However, as institutions increase their capacity to capture and analyze student information, they will likely need to address concerns regarding data privacy and establish a process for informing students of how their information will be used.

As administrators develop data-informed interventions to address students' needs, it will be critical that such strategies are based on the experiences of all students. For example, administrators that intend to use predictive models for the purpose of identifying at-risk students will need to be careful to avoid using engagement data in ways that lead to inherent bias, particularly with regard to identifying behavior patterns for underserved or underrepresented populations.

Predictive models are an attractive option for institutions that need a strategy for matching limited resources to students who are most in need. By including engagement and behavioral data in their models, institutions could strengthen the accuracy of their analyses and possibly increase the influence of support services on student retention. Several institutions in this study have had positive results from their application of predictive analytics, and other institutions similarly expect successful implementation in the next few years.

INTRODUCTION

As the cost for students to attend college continues to increase along with the cost to successfully operate higher education institutions, more national focus is being placed on the overall return on investment in higher education. For example, as of 2015, 32 states are using some form of performance funding for public colleges and universities, with student retention and persistence as a measure (National Conference of State Legislatures, 2015). Public scrutiny is high, and institutions must provide evidence of how their students are persisting toward a college degree. As a result, colleges and universities are expanding their use of data to improve student performance.

Several years ago, McKinsey & Company reported on the use of big data to heighten productivity across economic sectors and rated the education sector lower than most because of "a lack of data-driven mindset and available data" (Manyika et al., 2011, p. 9). However, the data environment within higher education has quickly shifted, as a wealth of data is now available and analyzed with varying degrees of sophistication (Dahlstrom, 2016). A number of institutions and technology vendors are responding to the call for data-driven decision making, and many are using predictive analytics.

> **Public scrutiny is high, and institutions must provide evidence of how their students are persisting toward a college degree.**

Predictive analytics is the "process of discovering, analyzing, and interpreting meaningful patterns from large amounts of data" (Patil, 2015, p. 138), a practice that has been widely used in business intelligence for decades. Higher education institutions have regularly used analytics to predict enrollment patterns for admissions and housing purposes, and it is now emerging as a strategy to improve student persistence and degree completion. In the context of student success in higher education, predictive analytics can be defined as the practice of collecting and analyzing student data to inform decision making regarding programs, services, and intervention strategies related to student persistence toward a college degree.

Well-documented successes at Georgia State University (GSU) provide examples of how strategic data collection, analysis, and targeted interventions can empower an institution to serve an increasingly diverse student body with typically extreme risk factors for attrition (Kurzweil & Wu, 2015). Over 10 years, GSU dramatically improved its 6-year graduation rate by 22% and increased its enrollment of traditionally underserved students by 27%. Improvement initiatives included several projects that utilized data analytics to determine achievement gaps and prescribe student-focused interventions to improve performance.

Institutions across the United States have observed GSU's success; as a result, many administrators also seek new ways to use student data and information to uncover factors that contribute to retention and graduation. Various institutional divisions and departments are collecting student information, with methods that range from simple paper forms to electronic methods such as swipes of student identification cards and tracking the rates at which students log in to the institution's network. Several vendors are helping institutions create, manage, and analyze large data sets, which often couple pre-enrollment variables with longitudinal academic records to show factors related to degree completion. If data patterns show common risk factors that lead to student attrition, institutions can address the factors through specific interventions to improve retention.

As technologies evolve and institutions expand their capacity to harness data, they will need even more cross-functional collaboration and efficient use of resources. In 2011, the Western Interstate Commission for Higher Education Cooperative for Educational Technologies founded the Predictive Analytics Reporting (PAR) Framework. This educational collaborative examines undergraduate student data from multiple institutions to find common trends at course, program, and institutional levels via a student success matrix, which maps strategically timed interventions to address attrition trends. The PAR Framework is useful for institutions that are not yet ready to analyze their own student data but want to apply national trend data to their work.

This study has identified a need for common language to describe predictive analytics. For instance, some higher education professionals perceive predictive analytics as the collection, analysis, and interpretation of data using statistical models, while others also include steps beyond interpretation of data—for example, employing data-informed strategies to achieve results. For the purpose of this study, predictive analytics was defined as "the collection and use of student data to inform decision making regarding programs, services, and intervention strategies related to students' persistence toward a college degree."

> **From creating common data definitions and shared student success intervention frameworks, our differences have become strengths. Each institution is openly sharing effective practices and major stumbling blocks with one another in the spirit of finding better pathways for encouraging student success.**
>
> ***—Ellen Wagner (2013), Chief Research and Strategy Officer, PAR Framework***

Although researchers are now examining the differences between institutional analytics and student success analytics, a deeper examination of how student affairs divisions contribute to institutionwide analytics efforts is needed. In terms of expanding the use of predictive analytics to include engagement and behavioral variables, many institutions already may have several of these data sets that could be utilized to inform retention efforts. Most student affairs departments collect data that can provide evidence of student engagement beyond classroom attendance and academic performance; however, such data may not yet be considered for predictive modeling as variables that significantly influence student retention.

Including student engagement and behavioral data in predictive algorithms can add a depth of understanding to help institutions more efficiently develop interventions to improve student success. For example, student engagement data include such measures as the frequency at which students visit campus departments, use support services, and participate in cocurricular programs. Student attendance data from functional units such as tutoring, advising, and career services can be connected to other data regarding students' connection to clubs and organizations, and then linked to the institution's student information system.

PURPOSE OF THE STUDY

Prior research has explored overall institutional use of predictive analytics, which in many instances primarily involves a focus on academic variables. The purpose of this study was to add to the existing literature by examining the level at which institutions also use student engagement data—particularly related to behaviors outside the classroom—in their predictive models. Much of these data are housed in the division of student affairs, which presents an opportunity to address how student affairs professionals collect and analyze data and their role in developing and implementing data-informed strategies to promote student success.

In addition to highlighting the institutional factors that influence the use of predictive analytics, this landscape analysis identified several challenges, opportunities, and considerations for utilizing student engagement data. The research also examined institution examples of data-informed interventions that are designed to increase student retention. The findings of this landscape analysis offer a foundation for deeper inquiry into the variables that are needed for robust predictive analytics models—that is, ones that can lead to the development of effective and efficient strategies for improving student outcomes.

METHODOLOGY

NASPA interviewed professionals from student affairs, academic affairs, institutional research, information technology, and assessment at 25 institutions via conference calls over a 6-week period in July and August 2016. Of the 25 institutions, 20 were public and 5 were private. Four of the institutions were community colleges. Participants described their current practices in collecting and using data, including direct connections to both student affairs initiatives and institutionwide analytics efforts. In addition, some participants described institution- or division-level interventions that were developed with results from predictive models.

DISCUSSION OF FINDINGS

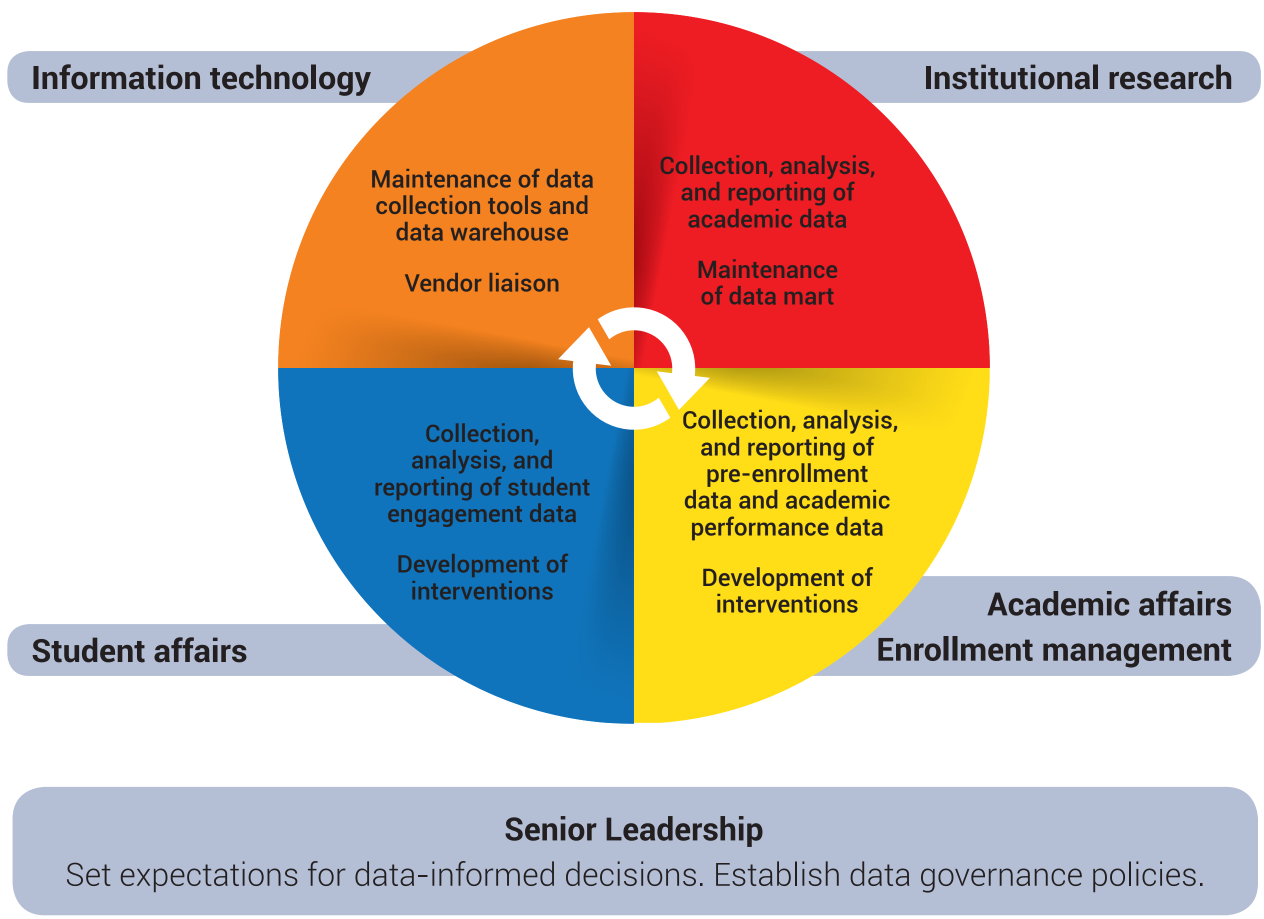
Institutions varied in their approach for implementing a predictive analytics initiative, primarily because each has a different set of institutional goals and different financial and personnel capacity levels. However, discussions revealed four common elements that describe such efforts: strong senior-level leadership of a cross-functional team approach; an institutionwide strategy for collecting, connecting, and accessing data from multiple systems; assessment of real-time response mechanisms; and ongoing communication and training.

Senior-level administrators are leading an institutional culture of data-informed decision making. Retention is a mission-critical issue for most institutions. In fact, nearly all respondents reported that a commitment to leveraging data for student success was evident in their institution's mission statement and strategic plan and was emphasized publicly by senior leadership. It is commonplace for senior leaders to expect student success data to be included in reports, discussed in meetings, and included in planning. For example, at several institutions, resource allocations are based on data that show areas of greatest student need.

Most institutions are planning their use of predictive analytics with a committee that is focused on enrollment management or retention. Typically, these committees include faculty and staff from multiple areas, including academic affairs, student affairs, institutional research (IR), and information technology (IT). Strong partnerships are formed between these functional areas, as such relationships are key to successful data analytics initiatives at both the institution and the division level. As academic affairs and student affairs divisions assume responsibility for managing certain elements of the institution's student success strategies, it is critical for staff from these units to effectively liaise with IR and IT colleagues to access and collect the necessary data variables for analysis.

Institutions that are working with a vendor sometimes include a vendor representative or campus liaison to the vendor on the committee. Institutions that have more established analytics systems continue to examine possible data variables to include in future predictive models. The core responsibilities for the primary committee members are shown in Figure 1.

Figure 1. **Typical Institutional Structures for Data Analytics Projects**



Data are collected and managed strategically for the purpose of improving enrollment management and increasing undergraduate retention. Some institutions are using predictive analytics for recruiting purposes, with a goal of building a successful student body. Institutions first aimed to use data to recruit students who were likely to succeed in college. That effort has progressed to using data to match enrolled students with the resources they need to persist.

> **Over the last 18–24 months it [the use of data analytics] has continued to evolve. The chief information officer and chief academic officer make this a high priority. Engagement data [are] important because enrollment must grow. Significant investment has been put toward it. How do we get to know the factors that contribute to the student experience? What are the data points that contribute to decisions? Use of data is modeled at the highest possible level.**
>
> ***—Michael Christakis, Vice President for Student Affairs, University at Albany***

Institutions reported that prior to considering the use of predictive analytics, they created data governance committees to make decisions about how data would be accessed, collected, analyzed, and reported across departments and divisions. These institutions also use a data warehouse, which is typically managed by the IT department as a central repository for core variables. The data warehouse typically stores and connects data from several systems, including the student information system, student organization database, housing software, building access records, learning management system, and academic support software.

It is common for individual users to access data from the warehouse via a tool often referred to as a data mart. The data mart provides faster access to frequently used data points, as faculty and staff can query information for their specific areas with little or no additional involvement from the IT or IR office. For example, if an institution's data warehouse contains advising, academic course, identification card, and student activities variables, a faculty or staff member could possibly run queries on students' participation in a club or organization, attendance at academic advising appointments, visits to the library, or use of the learning management system.

Institutions are assessing the effectiveness of their current programs and procedures, and preparing to adjust or reallocate resources if necessary. Before using predictive analytics to develop interventions, institutions are determining their level of readiness to respond to the findings of their analyses. As a result, some institutions are creating new staff roles or reorganizing existing ones to execute student success strategies. Institutions are also developing new programs or practices, adjusting existing support procedures, or discontinuing ineffective programs. For example, one institution found that despite the use of an early alert survey for several years, its overall retention rate did not improve. Professionals at the institution attributed the failure of the program to inadequate allocation of resources to effectively respond to the information from the early alert survey. Although the survey results were accurate, the delivery of interventions was inconsistent.

IVY TECH COMMUNITY COLLEGE has developed "NewT," an interactive platform to share data with faculty and staff on multiple campuses. Extensive training is being provided before its launch, and resources will continue to be available after implementation.

Communication is key. With this campuswide effort, there are messages that are now part of the institutional culture. We have to keep the communication going so people don't forget.

—Jason Meriwether, Vice Chancellor for Enrollment Management and Student Affairs, Indiana University Southeast

Ongoing communication and training is critical. As institutions employ more sophisticated data systems, users at all levels will need additional training on how to access information and use it to make decisions and develop interventions. For example, many senior leaders are now expecting staff from across the institution to create dashboards of useful information as part of a campuswide data analytics effort. Therefore, training on specific skills such as creating visual displays and on specific issues such as addressing data security and integrity will be important as more functional users have access to students' information. It will also be critical for departments to communicate broadly about the data they can provide for predictive models and student support strategies.

LEVELS OF ADOPTION

More than half of the respondents shared that they are still developing a process for using predictive analytics to support their retention efforts. Based on interview responses, three distinct levels of adoption emerged: planning, early implementation, and established practice (see Figure 2).

Figure 2. **Institutions' Use of Predictive Analytics for Retention Initiatives**

Planning. Five institutions indicated that they are in the planning phase and are preparing to launch their predictive analytics effort within the next year. Institutions in this phase are typically engaging in several core activities, including identifying the data variables that will be included in the predictive model, establishing a process for collecting and warehousing the data, and determining which personnel will be primarily responsible for analyzing the data and leading the development of interventions based on the results. All institutions in this phase reported having a cross-divisional planning group that is either specifically assigned to the predictive analytics effort or part of existing retention, advising, or enrollment management committees.

Personnel involved in the planning phase include senior-level administrators and staff with management responsibilities from several functional areas, including IR, IT, academic affairs, and student affairs. Several core student services functions—such as enrollment management, admissions, academic advising, career advising, and tutoring—are represented. In some instances, individuals with expertise in data management and statistical analysis are engaged in planning. As mentioned earlier, institutions that intend to work with a vendor also typically have a vendor representative or staff liaison as a member of the group.

> **We have great student support services, but we needed to collaborate and connect them in a way that would work. The program focused on students in the second and third tier because they were the ones being lost the most.**
> ***—Jason Meriwether, Vice Chancellor for Enrollment Management and Student Affairs, Indiana University Southeast***

Early implementation. Nine institutions were in the early implementation phase of using student success analytics. Many of these institutions have at least partially implemented a data analytics system for 2 to 4 years. These institutions are still refining their processes, either through their work with a vendor or adjustments to their homegrown method. Of the 14 institutions in the aforementioned planning phase or the early implementation phase, 7 are working with a vendor, 2 are looking for a vendor, and 5 are utilizing in-house data analytics tools. In some cases, senior student affairs leaders are responsible for the effort, but it is more often led by the IR or institutional effectiveness office. While institutional committees are usually larger and more active during the planning phase and initial launch of the effort, a smaller core group of key administrators continue the work after implementation.

Established practice. Eleven institutions had established data analytics practices. Several of these institutions reported that the process began as a way to create a successful student profile for admissions recruiting purposes and evolved into an analysis of factors related to student retention. These institutions have been engaged in their efforts for at least 5 years. As a result,

CALIFORNIA STATE UNIVERSITY, CHANNEL ISLANDS, has convened an institutional data governance committee to ensure that the multiple demands for collecting and analyzing data are prioritized. The committee is focusing on where data exist, the time frame for collection, and how data are described so they are clearly accessible to those who use the data sets.

VALDOSTA STATE UNIVERSITY mapped the student lifecycle from registration to enrollment, to tutoring, to advising, and looked for ways to improve processes and direct resources to help retain students.

the use of data analytics is embedded into the work of retention and advising committees, and information is shared with senior leadership. Colleges and universities with established analytics processes continue to refine their work to make data more accessible to all areas of the institution. Recent innovations in technology, as well as changing student populations and campus environments, has led to continued evolution of data analytics work. Five of the established institutions use existing data platforms or specially designed in-house systems to gather data and run predictive algorithms.

Some institutions with sophisticated in-house resources for data analytics still see an advantage of working with a vendor. For example, some vendors may help an institution merge data from multiple sources and systems across the campus, which could help the institution improve the quality of its student-level analyses and better address state performance metrics and retention goals. Three institutions are working with a predictive analytics vendor, and three are in the process of contracting with a vendor to enhance their in-house efforts.

TYPES OF STUDENT SUCCESS DATA

Although institutions collect countless amounts of data every day, the task of determining how to use the information can be overwhelming. Figure 3 displays five categories of student success data and possible data variables for collection that emerged from discussions with the institutions: **pre-enrollment** data, which include demographic characteristics, test scores, and prior academic performance; **academic** data, which address student participation and performance in class; **motivation and self-efficacy** data, which relate to a student's ability to adjust to the campus culture; **use of support services** data, which pertain to functions that are designed to help a student succeed; and **student engagement** data, which indicate a student's level of integration with the institution, both socially and in cocurricular environments. The formats of data collection are varied, but the scope of information is the same in that most predictive models begin and end with pre-enrollment and academic data. Other models also use motivation and self-efficacy data gained from first-semester check-in surveys to measure levels of each that can be flagged in a student's record.

INDIANA UNIVERSITY SOUTHEAST initially assumed that students were leaving due to transfer to nearby campuses. However, they examined student performance over time—specifically retention and graduation over 8 semesters, 4 years, and 6 years—and discovered that their assumption was inaccurate. Their analysis showed that they were losing large numbers of "B" students, who were middle performers not currently receiving intervention. Further analysis of students who departed the institution showed students were leaving because of life circumstances, rather than transfer to other institutions.

INDIANA UNIVERSITY SOUTHEAST increased its outreach efforts and captured through its student information system the number of "touches" a student experienced, such as academic advising, financial aid, and housing staff. Faculty were asked to have an extra conversation with students, and department heads and deans also engaged in extra outreach. In the first attempt at this targeted intervention, the institution increased its fall to spring retention by 4%.

Figure 3. **Types of Student Success Data**

Pre-enrollment	Academic	Motivation and Self-efficacy	Use of Support Services	Student Engagement
• Demographics • High school grade point average • Parents' experience with college • Test scores	• Class attendance • First semester grades • Grades in select core courses • Login to student web portal • Midterm grades • Registration for next semester • Use of learning management system	• Comfort with academic ability • Depression • Financial issues • Homesickness • Lack of friends or connections	• Advising • Career services • Counseling • Disability support • Financial aid • Health center • Library • Tutoring	• Athletic team affiliation • Campus membership • Campus residency • Campus Wi-Fi usage • Dining center • Leadership roles • Participation in campus programs • Recreation center

CONSIDERATIONS

"Enterprise systems aren't really set up to collect student affairs data. We have to come up with creative ways to collect and share data so engagement data can be utilized.

—Tyneka Harris Coronado, Project Leader—Student Affairs, DePaul University"

While student engagement and behavioral data may be available, there appear to be several reasons why they are not used in predictive models. One main obstacle is a lack of data collection methods that provide consistent information over time and that can be connected to other systems. Pre-enrollment data and most academic data are already housed in the institution's student information system. Data from learning management systems can be imported into the student information system; however, other data collections cannot. In some instances, the data collection process for such valuable information as students' use of student services and their engagement with the institution is still paper based. In other cases, institutions are using platforms that allow student engagement information to be tracked, but such systems often do not connect directly to the student information system. Of the listed types of data for the Use of Support Services category, advising and tutoring are the most likely to be included in a predictive model because use of those services is often tracked in student information systems.

The use of data from swipes of a student identification card may sound simple, but this data collection method presents several obstacles. The hardware needed to collect data with this method is cost-prohibitive for many departments and is somewhat impractical to use at events that several hundred students attend. However, some institutions are making strong progress with collecting identification card data.

Conversations with institutions also revealed that data provided from student affairs departments are sometimes "messy" in terms of

OREGON STATE UNIVERSITY (OSU) is leading an initiative in student affairs to make card swipe technology available to all student affairs units. Card swiping is typically tracked for housing and campus recreation departments because the technology is already being used for entry purposes, and evidence of the value of these entities on campus has been studied in recent years. Last year, OSU collected 78,000 additional points of contact that were not historically available. All student affairs programs are now connected to five learning domains, and participant data can be tracked over time.

"Student affairs and academic affairs already have an excellent connection here. The beauty of this system is there will be coordinating and sharing of data across academic affairs and student affairs areas that can be accessed by those helping a particular student."

—Dennis Pruitt, Vice President for Student Affairs, University of South Carolina

accuracy and consistency of collection over time. For example, one institution reported that several of its student involvement programs collect participant information by name only and do not collect a student identification number. The institution is attempting to collect data for a longitudinal analysis, but when matching student participation to academic persistence, professionals are sometimes unsure about whether student data are matched accurately.

IDENTIFYING STUDENTS FOR INTERVENTION

An impetus for using predictive analytics is to apply limited resources to intervene with the students who are likely to be retained with extra support. While support services are generally available to all students at an institution, targeted outreach can ensure that a student who is in danger of leaving the institution will be referred to the appropriate resources and remain enrolled. Each institution that participated in this study had different retention goals, with most efforts focused on retaining first-year, first-time college students. Other institutions are expanding their retention initiatives to later in the student lifecycle to see persistence gains. There are differing approaches to identifying cohorts of students using predictive analytics, but institutions appear to be gearing their efforts toward three primary student groups:

Students who need high-touch intervention – Intrusive advising and support is implemented for students who are flagged with several risk factors such as low class attendance, poor academic performance, or survey responses

At the **UNIVERSITY OF SOUTH CAROLINA,** a major effort is underway to connect specific data points of student engagement to their predictive analytics models. The university's Beyond the Classroom Matters initiative is cataloging specific student leadership and involvement experiences with documented learning outcomes for the primary purpose of creating a comprehensive academic record that includes cocurricular experiences.

that indicate low engagement with the institution. Data are used to identify manageable groups that can receive support.

Students who need moderate-touch intervention – A predictive analytics product is designed to help institutions reach students who earned a 2.0–3.0 grade point average in their first year (Tyson, 2014). This approach identifies students who did well enough not to require an intervention during their first year but are more at risk after the first year. These students are reported to be likely to succeed with targeted support.

Students who need immediate intervention – Some institutions have identified specific, easily accessible data points that can trigger attention for students who are likely to leave. For example, three institutions shared that students who do not swipe their identification card at any campus program or service in the first 6 weeks of the semester are flagged for outreach. At one institution, students who are doing well academically but request an academic transcript from the registrar will get a personal call from an advisor. A transcript request could be an indicator that the student intends to leave the institution so an advisor contacts the student to inquire about the kind of support they might need in order to not leave the institution.

> “**Long-term, we hope higher education data management products will integrate. Just as the Microsoft Office Suite was a bunch of stand-alone products, it would be easier if all these student data platforms were integrated. Academic advising should be integrated with degree audit, which should integrate with academic holds, and putting all the different parts together. But until then, the institution has to function as if all the data are integrated, which creates more work.**
> ***—Nolan Davis, Chief Student Affairs Officer, Elizabeth City State University***”

Student engagement data are also used to gauge success of interventions. Institutions can track referrals of students to various services and then examine whether students actually use the service. Many institutions embed alert flags and follow-up notes into an online advising system that can be accessed by any faculty or staff members assigned to the student. The aim of all this work is to help students persist to graduation—and it appears

> **These are the students we want to keep, rather than have them transfer to another institution.**
> ***—Jeffrey Hoyt, Assistant Vice Provost for Institutional Effectiveness and Analysis, Florida Atlantic University***

effective, as one institution examined data for performance metrics and realized that retaining an additional 35 students represented a full percentage point on their retention dashboard.

By using data analytics to focus interventions on students who would likely stay enrolled with extra attention, an institution can accomplish its retention goals more efficiently. However, when utilizing predictive algorithms, institutions will need to consider the possibility of inherent bias that may marginalize students from underrepresented populations. Data analysis is unusually susceptible to reproducing discrimination and privilege through researcher bias, confirmation bias, and an emphasis on iterative model generation (DeLuca Fernández & Newhart, 2016). This consideration is important as more institutions increase their capacity to collect data from multiple offices, especially those that serve underrepresented populations.

STUDENT PRIVACY

The increased use of data to help students succeed has led some to perceive that such usage has the potential to become invasive. Some institutions may question whether students actually want increased tracking of their engagement levels. One institution respondent shared that when students were asked their opinion about increased tracking of their participation in campus programs, the response was that they thought the institution was already doing so. The institution still posts a privacy notice at every event that addresses how data from students' identification cards are used. Some institutions are planning to mimic consumer applications of technology. Similar to how retail businesses text coupons to nearby customers, two institutions shared their intention to use Global Positioning System (GPS) technology in the future to track where their students are logged in to their Wi-Fi system.

For example, if the system detects a student near a library or an advising center, the student could receive a notification or text message encouraging the student to use the services or attend events that are nearby.

A different kind of GPS program is the Guided Pathways to Success, sponsored by Complete College America. The GPS Direct Initiative recently launched and offers a "Technology Seal of Approval" for institutions that incorporate intrusive advising practices into online advising platforms. One example of a tactic endorsed is a pre-populated course schedule provided to the student, with courses scheduled based on predictive data about the student and analytics related to course load and sequence (Complete College America, 2016). Administrators who are concerned with student privacy suggest that some students may object to a pre-filled schedule and would prefer to make their own choices. Institutions may need to consider the need to issue a privacy statement to students, similar to those issued by credit card and software companies.

> **We will need to inform students about the kind of information we collect about them and how we intend to use it, and we may need to allow them to opt out.**
> ***—Dennis Pruitt, Vice President for Student Affairs, University of South Carolina***

USE OF EARLY ALERT SYSTEMS

Although the majority of the institutions are not yet using engagement and behavioral data in their predictive models, it appears that the use of data for the purpose of early alert systems is especially prevalent. An early alert system identifies students for intervention within an academic term in order to encourage student success and retention. Whether the alert is manual, such as a faculty member sending an e-mail to the administrator in charge of student support, or embedded in the student information system, such as online triggers that are sent to designated personnel who will intervene, the basis of early alert systems is the use of monitored student data points that trigger an action to be taken. Nearly all of the participating institutions

(23 of 25) are either developing or enhancing this type of system. The two institutions that were not yet using an early alert system reported that they plan to do so in the future.

Early alert systems are generally designed for new or transfer students in their first semester; however, some institutions reported an increased interest in using some data points such as class attendance and use of the learning management system for students at all levels. Once the early alert system is put into action, designated faculty, staff, or student peer mentors are deployed to intervene with students. For example, if an institution observes a lack of data regarding a student's use of support services or limited data related to student engagement, the early alert system may generate a warning. The warning may lead to intentional interventions from staff in student affairs or other divisions. Such interventions can be in the form of a check-in phone call or visit with the student, a targeted e-mail with information about support services and opportunities, scheduled advising appointments, or other forms of contact. Depending on the structure of the institution, these efforts are coordinated by academic affairs or student affairs, and are sometimes a collaborative effort between departments and divisions.

In a strict interpretation of predictive analytics, an institution will use its own longitudinal data to determine factors related to retention and persistence, and these factors will be used as the behavioral data points in an early alert system. However, not all available student data are currently included in institutional predictive models, even at institutions with established predictive analytics practices. If there is a theoretical basis for looking at a particular student behavior as a factor in retention, an institution can still track real-time student behavior on that point and utilize it as an early alert for intervention. For example, several institutions cited campus results from national student engagement surveys as the basis for their early alert triggers.

The same data points that can be used in a predictive student success model can be used as trigger points for early alert systems on a student-by-student basis. With student engagement data, institutions can often use information that is already being collected for program assessment purposes.

One institution with limited resources is still finding ways to identify students for intervention using data analytics. The institution developed a database to analyze each student on 10 risk factors, including pre-enrollment data, financial aid, and student engagement information such as athletic team affiliation and residency status. Each factor was rated on a 3-point scale (no problem, needs attention, and critical). The institutional research office then developed a predictive model for retention. Combining this information with a campuswide early alert network resulted in a team of administrators monitoring student progress and students being identified for intervention by the dean of students.

> **The big adjustment from student affairs assessment to early alert interventions is the real-time urgency. If you want to help students, you have to collect and analyze the data quickly, so you can act upon the data to intervene.**
> ***—Michael Christakis, Vice President for Student Affairs, University at Albany***

All institutions that reported having an early alert system are using pre-enrollment and academic data (see Figure 3) as intervention triggers. All systems involve a reporting process in which faculty indicate concern based on such data variables as class attendance or early-term grades. This information is either reported manually by faculty members or tracked in a learning management system. Eight institutions reported using check-in surveys, which measure a student's motivation and self-efficacy a few weeks after the start of an academic term. Information from the Use of Support Services and Student Engagement categories is used as triggers by 10 institutions. Five additional institutions reported plans to use some form of this engagement information in the future.

CONCLUSION

Most institutions are using data analytics to some extent, whether to monitor operational efficiencies or to inform student success initiatives. Institutions have the potential to analyze data points from nearly every interaction they have with a student. As technology evolves, so will the capacity for higher education professionals to serve students in ways that will better help them to persist and earn a college degree. This landscape analysis examined the use of predictive analytics and the extent to which predictive models included not only academic variables but also student engagement and behavioral data. Interviews with the 25 participating institutions revealed that there are opportunities to further expand the use of these data, especially because the information is often already collected for other purposes. It appears that many institutions are considering using more of these data and have established strong cross-functional collaborations. However, these teams may not be ready to fully implement a robust predictive analytics system that includes a wider range of engagement variables for another couple of years, as they are still addressing challenges with connecting data from sources external to the student information system.

This study offers a foundation for additional research on institutions' use of predictive analytics, particularly with a focus on deeper exploration of robust student success analytics systems. Additional research of these systems would provide valuable insight regarding the optimal level of data collection, including the number and types of variables that should be collected, the specific personnel and financial capacities needed to sustain the effort, the types of interventions that are successfully developed and implemented based on data analytics, and connections between engagement data and noncognitive variables such as communication, leadership, and self-efficacy.

NASPA will use the themes from this initial review to continue examining how institutions use predictive analyses of engagement and behavioral data. Next steps for this research include a national survey to uncover differences by size, sector, and student population. Of particular focus will be the methods by which institutions successfully integrate engagement and behavioral data into their student information system, and common challenges that occur during the process. For example, this study confirmed that it is critical for institutions to have high levels of coordination between the many units that collect and analyze student data. As a result, it is also imperative that multiple professionals across the institution have the capacity to interpret the results of predictive models and inform decisions regarding the appropriate time and method for applying interventions to improve student performance.

The forthcoming national survey will address how institutions measure the effectiveness of their predictive models and select interventions on students' retention and graduation. For example, the survey will examine the types of academic, engagement, and behavioral data variables that contribute to strong predictive models. One benefit of the national survey is that it will be disseminated to institutions of all sizes and sectors, which should result in a more precise discussion of similarities and differences in how institutions build data capacity across departments, functions, and divisions, as well as how they use predictive models to develop interventions.

"The forthcoming national survey will address how institutions measure the effectiveness of their predictive models and select interventions on students' retention and graduation."

The institutions in this study confirmed that the goal of retaining and graduating students is the primary driver of the decision to use predictive models. As we continue to examine the additional factors that influence the use of predictive analytics, we must also suggest strategies for how institutions can strengthen and sustain a culture of evidence-based decision making. As institutions increase their capacity to use sophisticated data models, it will be especially important for the higher education community to share knowledge and resources to help professionals effectively address data governance and data privacy, both of which will be areas for ongoing discussion in the years ahead.

REFERENCES

Arroway, P., Morgan, M., O'Keefe, M., & Yanosky, R. (2016). *Learning analytics in higher education*. Retrieved from Educause Center for Analysis and Research website: https://library.educause.edu/~/media/files/library/2/2016/ers1504

Complete College America. (2016). *Complete College America unveils technology seal of approval* [Press release]. Retrieved from http://completecollege.org/complete-college-america-unveils-technology-seal-of-approval

Dahlstrom, E. (2016, September/October). Moving the Red Queen forward: Maturing analytics capabilities in higher education. *Educause Review*, 36–54.

DeLuca Fernández, S., & Newhart, D. (2016, March). *Assessment and the seduction of big data: How to avoid big mistakes and promote equity*. Presentation at the 2016 NASPA Annual Conference, Indianapolis, IN.

Kurzweil, M., & Wu, D. D. (2015). *Building a path to student success at Georgia State University*. Retrieved from ITHAKA S&R website: http://www.sr.ithaka.org/wp-content/uploads/2015/08/SR_Case_Study_Building_Pathway_Student_Success_042315_0.pdf

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition and productivity*. Retrieved from McKinsey Global Institute website: http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation

National Conference of State Legislatures. (2015, July 31). *Performance-based funding for higher education*. Retrieved from http://www.ncsl.org/research/education/performance-funding.aspx

Patil, S. (2015). Predictive analytics in higher education. *International Journal of Advanced Research in Computer and Communication Engineering*, *4*(12), 138–141. doi: 10.17148/IJARCCE.2015.4123

Tyson, C. (2014, September 10). The murky middle. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2014/09/10/maximize-graduation-rates-colleges-should-focus-middle-range-students-research-shows

Wagner, E. (2013, December). The predictive analytics reporting (PAR) framework, WCET. *Educause Review*. Retrieved from http://er.educause.edu/articles/2013/12 the-predictive-analytics-reporting-par-framework-wcet

Yanosky, R., & Arroway, P. (2015) *The analytics landscape in higher education, 2015*. Retrieved from Educause Center for Analysis and Research website: https://library.educause.edu/~/media/files/library/2015/5/ers1504cl.pdf

APPENDIX A: METHODOLOGY

NASPA invited a wide range of institutions to participate in a research study on the use of predictive analytics. This convenience sample included senior student affairs officers, institutional research directors, and student affairs professionals engaged in assessment from 25 colleges and universities. Participants were interviewed via conference calls over a 6-week period in July and August 2016. In several calls, collaborative colleagues from academic affairs, institutional effectiveness, and information technology also participated. The result is a snapshot of a range of methods and approaches to utilizing data analytics to inform retention efforts. Twenty public institutions (including four community colleges and sixteen 4-year institutions) and five private, 4-year institutions are represented in the study. The institutions differ in size and scope, but all are committed to student retention efforts and are using (or planning to use) some form of data analytics related to student success.

Using a student affairs perspective, participants were asked to describe their current practices in collecting and using data, both in the context of student affairs work and with institutionwide analytics efforts. The aim of this study is to increase the understanding of how student engagement data can be considered as a factor in student retention, and how student engagement information can fit into the mix of data used for predictive analytics. In addition, participants described the interventions that are developed using the results of the predictive model.

The interviews were conducted using the following guiding questions and requests:

1. Provide a brief overview of how your institution currently utilizes predictive analytics and any additional efforts planned for the future.
2. Which student affairs functional departments utilize predictive analytics

to make decisions about programs and services? How do the departments use the data? How are staff assigned to the work?

3. In addition to student affairs, which campus personnel, if any, are involved in predictive analysis of engagement data? And how? Please define typical roles.
4. From your perspective, at what capacity does leadership at your institution use predictive analytics for decision making (within student affairs and the larger campus)? What factors are in place to support the use of predictive analytics?
5. What sources of engagement data are being used or developed that can or will aid in predictive analytics? Please describe places you would like to use it (paper vs. technology). Is data collection a barrier?
6. Is there data available that is NOT being used for predictive analytics? What are the reasons for not using this data?
7. What technology is used for data collection and analysis? Is this technology product from a contracted vendor, an in-house system, or a combination of both? Which campus personnel know how to use the technology?
8. What are some specific examples of how engagement data are being utilized on your campus to inform student retention efforts?

Transcripts of interviews were reviewed for thematic patterns in responses. Findings within these themes are reported in the aggregate. Specific examples of current practices that were discussed in the interviews are included in this report. The information gathered through this study can inform future research and recommendations about the use of student engagement data in predictive analytics.

APPENDIX B: PARTICIPATING INSTITUTIONS

INSTITUTION	ENROLLMENT
Austin Peay State University	10,000 to 14,999
California State University, Fresno	20,000 to 29,999
California State University, Channel Islands	5,000 to 9,999
Chatham University	Less than 5,000
DePaul University	20,000 to 29,999
El Paso Community College	20,000 to 29,999
Elizabeth City State University	Less than 5,000
Florida Atlantic University	30,000+
Grand Valley State University	20,000 to 29,999
Indiana University Southeast	5,000 to 9,999
Ivy Tech Community College	30,000+
Montgomery County Community College	10,000 to 14,999
Oregon State University	20,000 to 29,999
St. John's University	Less than 5,000
Saint Louis University	15,000 to 19,999
Stephen F. Austin State University	10,000 to 14,999
Stonehill College	Less than 5,000
University at Albany, State University of New York	15,000 to 19,999
Texarkana College	Less than 5,000
University of Central Oklahoma	15,000 to 19,999
University of New Mexico	20,000 to 29,999
University of North Carolina Wilmington	10,000 to 14,999
University of South Carolina	30,000+
University of Utah	30,000+
Valdosta State University	10,000 to 14,999

APPENDIX C: TECHNOLOGY TOOLS UTILIZED

Note: NASPA does not formally endorse any of the vendors or service providers listed here. Institutions that participated in the study were not selected based on their use or nonuse of any of the following tools.

Predictive Analytics

- Education Advisory Board/ Student Success Collaborative
- Civitas/Illume
- SAS (in-house data analysis)

Student Information Systems

- Banner
- PeopleSoft
- Jenzabar

Data Warehousing

- Oracle
- SQL Server
- Amazon Web Services

Dashboards and Data Marts

- Tableau

Learning Management Systems

- BlackBoard
- D2L
- Jenzabar

Advising Tools to Flag Students and Track Advising Interventions

- GradesFirst
- Skyfactor
- Starfish
- TutorTrac

Student Engagement Data Tracking Tools

- OrgSync/Campus Labs
- Presence
- Pocket Tracker (Blackboard)
- Symplicity (Career Services)

NASPA® | Student Affairs Administrators in Higher Education