



# What would you do with all this data?

Promoting Student Success with Predictive Analytics

# Defining Analytics for Higher Education

---

Predictive analytics in higher education has grown markedly in recent years, following the trend of using data to drive decisions in all industries – both public and private. While a well-defined and accepted part of the lexicon among the business and technology industries, predictive analytics in higher education is still a relatively nascent idea, and as such a common vocabulary and understanding is still being developed.

The main idea behind predictive analytics is twofold: mining large sets of information for insight into patterns of behavior, and using that insight to drive decision making. Businesses have been identifying many ways to harness the power of predictive analytics. For example: Target mines the aggregate data of its debit and credit card users' purchases, and offers individualized coupons for future purchases based on purchase history. This encourages the shopper to return to Target, with coupons specific to his/her needs. Netflix provides TV and movie recommendations based on subscribers' provided preferences and previously viewed and rated titles. And perhaps most famously, Google Analytics is an entire suite of data mining services designed to provide tailored ads to internet browsers based on their search habits and website visits.

"Big data" refers to analysis of datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. With the ever-increasing use of technology – including the internet, learning management systems, and social media – to gather student and institutional data, there is a growing electronic trail of information that can be mined for insights into student behaviors and indicators. Dates of registration, amount of time spent on task in an online course, scanning of student ID cards, and popular course sequencing (pre-requisites) are examples of interactions that leave a digital footprint, and can be used to draw conclusions about the higher education process and ways to make improvements. Specifically, analysis of this student data can provide insight into at-risk behaviors and indications where additional support – both academic and non-academic – may be needed.

EDUCAUSE, a nonprofit association whose mission is to advance higher education by promoting the intelligent use of information technology, has been instrumental in fostering discourse on predictive analytics by hosting conferences and webinars that gather data scientists with higher education technology professionals, and publishing white papers and also research papers on the use of analytics in higher education. EDUCAUSE has developed the following working definition of analytics in the context of higher education:

---

Analytics is the use of data, statistical analysis, and explanatory and predictive models to gain insights and act on complex issues.

This definition emphasizes the role of analytics in the overall process of strategic planning and decision making – an important distinction for higher education professionals. Analytics is a means by which administrators, faculty, and staff can understand their students and make informed decisions about their actions. As noted by Linda Baer and Donald Norris, analytics and big data offer the potential to identify promising practices, effective and efficient models, and powerful innovations, sustaining higher education for the future.

In the book, *Game Changers: Education and Information Technologies*, Vice Provost Vernon Smith wrote, “... higher education would use the power of information technology to mine student information and data on a massive scale across multiple institutions.... The idea would be to then create predictive models through the use of advanced statistical modeling that would identify possible stumbling blocks and help drive early interventions for students, especially low-income young adults and minorities. A growing body of best practices and interventions that remove barriers to student progress and success exists, but those interventions would be better informed if they were based on what the research and actual behaviors indicate, rather than on anecdotal notions or experience alone.”

The best of the leading higher education institutions are already embedding predictive analytics into both academic and administrative processes. These leading-edge practitioners have demonstrated that analytics can be leveraged to effectively support student success practices in a wide range of institutional settings. Many schools use predictive analytics so that at-risk behavior thresholds can be established as tripwires that provoke automatic, yet tailored, interventions, depending on the students’ characteristics.

To further demonstrate the benefits of utilizing analytics in higher education, the article, “Penetrating the Fog: Analytics in Learning and Education,” describes some of the following components when analytics is employed:

- Create, through transparent data and analysis, a shared understanding of the institution’s successes and challenges.
- Help leaders transition to holistic decision making through analyses of what-if scenarios and experimentation to explore how various elements within a complex discipline (e.g., retaining students, reducing costs) connect and the impact of changing core elements.
- Improve administrative decision making and organizational resource allocation.
- Identify at-risk learners and provide intervention to assist learners in achieving success.
- Provide learners with insight into their own learning habits and give recommendations for improvement.

- Innovate and transform the college/university system, as well as academic models and pedagogical approaches.

It is becoming clear that institutions that are able to embrace and adopt cultures centered around utilizing data in new and innovate ways will experience high levels of improvement in student success, recruitment, and retention.

## Practices and Perceptions of TBR Institutions

---

The institutions within the TBR system and the State as a whole are in a very viable analytics position of having access to the kind of large, aggregated datasets necessary to perform meaningful analysis and to be able to adequately train predictive models for improved accuracy. However, harnessing the power of this position will take a concerted effort both at the institutional and stat level. Organizational capacity for the use of analytics requires significant investment in human resources and technology capital, and must be a strategic priority for leadership and key stakeholders at each institution.

Understanding the data culture currently at TBR universities and community colleges is the first step necessary in moving forward as a system to truly capture increases in student success as afforded by effective use of analytics.

For the purposes of this study, TBR's 19 universities and community colleges were divided into two groups: the institutions that opted-in to participation in the Trajectory Analytics pilot, and the institutions that did not. More details about the pilot can be found in the section below.

The chief academic officers of the pilot institutions received a survey asking about their planned use of the pilot data, and also asking for an assessment of the institution's capacity for working with big data. Institutions not involved with the pilot were asked about their awareness and perception of the pilot data, and also for an assessment of their institution's capacity for working with big data.

Survey Response Rates

	Community Colleges	Universities
Pilot Schools	7/8 = 87.5%	2/5 = 40%
Non-Pilot Schools	3/5 = 60%	1/1 = 100%

With responses from a total of 13 out of 19 institutions, the sections following report the qualitative responses for both groups of institutions and provides an analysis of the state of data usage overall at TBR community colleges and universities.

## Trajectory Analytics Pilot

The Trajectory Analytics report, alternately called the Intrusive Advising or Intervention Analytics report/dataset, is an attempt at the system-wide level, to provide an analytics tool to institutions in order to develop strategies that increase student success through retention and persistence to graduation. It is a predictive model which, by analyzing historical student data aggregated from across the system, provides a measure of the student's progress towards completion of his/her selected program (their trajectory), and an empirical estimate of the probability of graduation, or completion, based on the student's academic history and trajectory. For each institution, students who had flat or downhill trajectories with very low probability of graduation were marked for intervention and collected into a spreadsheet data file.

During the spring semester of 2015, the datasets of students recommended for intervention were made available to all TBR universities and community colleges. Schools that opted in for participation were provided an Excel file and data codebook explaining each of the variables in the dataset. Some of those variables included:

- THEC CIP code and name of the student's current major (when available),
- Number of semesters enrolled and number of grades earned at the institution,
- A graphic depicting the trajectory of the student over time,
- The estimated graduation probability,
- Numbers of each letter grade, including critical grades in major fingerprint or graduation fingerprint courses (explained below),
- Age, high school GPA and ACT subscores for Math, Science, English, and Reading (where applicable),
- Full-time/part-time enrollment status.

This first summary page was followed by a spreadsheet for the annotated transcript of each student. The transcript spreadsheet listed each of the classes the student had taken, the grade obtained, whether it was a critical grade in a fingerprint course, and the course instructional delivery method.

Additionally, the report generated a list of courses at the institution that play a significant curricular role in their student success – called fingerprint and obstacle courses. These courses have been found to be disproportionately critical to overall success towards completion. A fingerprint course (for graduation or for a major) is course that historically students who have completed a degree program have been successful in – and students who did not graduate were unsuccessful in that course. Obstacle courses, and corresponding grades in that course, are those that most often cause student trajectories to become at risk.

One example of graduation fingerprint course at Pellissippi State is ENGL 1010, with a grade of C. That is to say, that of students who successfully graduated in any degree program, most had grades in ENGL 1010 of C or better. Whereas graduation fingerprint courses apply to graduation in general, major fingerprint courses are unique to a particular degree or certificate program. LAW 1000 with a grade of B and EDU 2010 with a grade of A are major fingerprint courses for the Paralegal Studies AAS and Teaching AST majors at Pellissippi State, respectively. An example of an obstacle course/grade is an F in HIST 1010. Students with this course grade typically exhibit low overall progress (trajectory) towards graduation.

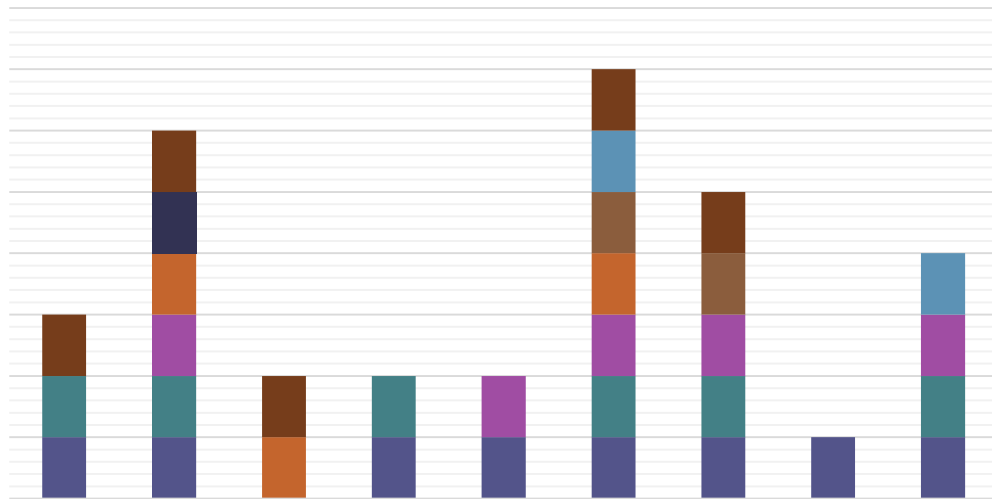
As of August 2015, eight community colleges and five universities received the intervention dataset (referred to as pilot schools). Those schools were encouraged to use the data report to independently develop institution-specific intervention strategies for these academically at-risk students.

## INSTITUTIONAL INVOLVEMENT

In “Building Organizational Capacity for Analytics,” a publication of EDUCAUSE, the authors Linda Baer and Donald Norris stress that successful implementation and strategy building from insights gained through analytics requires a cross-institutional team of Academic Affairs, Student Affairs and Enrollment Management, Institutional Research, Assessment and Planning, and Information Technology. The authors go on further to indicate the need for engagement of deans, department chairs, and administrative assistants in analytics applications. In fact, “the more people that touch the data, the better the understanding.”

Before delving into the planned strategies for the trajectory analytics data, we asked institutions, “*Who has been involved in exploring the trajectory analytics data?*” Results are in the chart below.

## Institutional Involvement (n=9)



Each column bar on the chart represents the reported offices involved at each responding institution. The “other” category included Institutional Effectiveness and Planning – distinct from the Office of Institutional Research (1), Information Technology (2), Associate VP of Faculty, Curriculum, and Programs (1), Academic Deans (1), Academic Assessment Specialist (1), Faculty member under special stipend (1), and Enrollment Management and Academic Support (1).

In over half of the institutions responding, fewer than four offices were involved in exploring the data and developing strategies. This points to a lack of institution-wide buy-in and understanding of interdisciplinary approach that is most effective with analytics strategies.

Also interesting to note: although the data was distributed through the chief academic officer (Vice President of Academic Affairs or Provost), all but one institution involved the office of Institutional Research – and only two institutions had members from the office of Information Technology involved.

## FEATURES OF THE DATA

In an effort to better serve institutions with future analytics reports and to elicit feedback on this intervention report, the survey also asked institutions to identify features of the trajectory analytics report found most and least helpful. Throughout the pilot process, the Central Academic Affairs Office worked to update the data file for readability and user-friendliness. Updates made during the pilot period included adding students' age, high school GPA, ACT subscores, and replacing some of the more complex statistics with graphical representations.

Institutions responded that the annotated transcripts provided for each flagged student provided a starting point to, "dig down into the student's academic journey." For academic advisors, this is invaluable, as it succinctly lists courses taken, and grade obtained for each student. The data also identify each student's major CIP code and number of enrolled semesters. Institutions reported that this information is also useful for starting conversations between advisors, faculty, and students.

The list of fingerprint and obstacle courses for the institution were also listed as a beneficial feature of the report. This information helped academic advisors identify specific barriers to completion in the various degrees for conversations with all students – not just those flagged as at-risk by the intervention analytics.

While the report focuses on the student's academic history, several institutions stated that financial aid information and student holds are important features that need to be added to the report. Understanding that academics is not the only factor in retention, one school noted, "if a student has run out of financial aid eligibility, a different conversation needs to be had other than pointing them to tutors."

Demographic information – specifically by subpopulations as outlined in the 2015-20 Quality Assurance Funding formula – was also requested. Knowing if a student is in a targeted demographic status allows the intervention strategy to funnel that student into existing initiatives, for example Veterans or adult students. Avoiding "initiative fatigue," or the over-creation of new strategies before fully realizing and/or capitalizing on existing ones, is an important thing for institutional leadership to keep in mind. Use of predictive analytics should enhance action plans already in place for increasing student success.

To facilitate that integration of the analytics report into existing institutional processes, one school noted the need for making the data available in other formats, such as SAS, that would require less upfront data manipulation.



## INTERVENTION STRATEGIES

The survey next asked the pilot institutions if any intervention strategies had been developed for the Fall 2015 semester using the trajectory analytics report, and to detail the target population and specifics of the strategies. Responses varied – four of the responding institutions did not have any plans for the Fall 2015 semester, while the other five provided intervention strategies ranging from general to very specific.

Each of the reported strategies attempted to target at-risk sectors of the institution's population. For some institutions, that meant looking at specific groups within a major or program, while other schools provided an intervention strategy for students who had not yet declared a major in their first year or who would benefit from a change in majors. As a result of recent initiatives in the State such as *Drive to 55* and the release of the new Quality Assurance Funding formula, some institutions targeted populations of high priority to those initiatives.

The intended intervention strategies did share the common theme of proactive advising and providing early alerts of at-risk student behaviors. Advising strategies ranged from providing academic advisors with information (fingerprint courses, obstacle courses) from the analytics report to mandatory advisement with faculty advisors and success coaches to develop academic, career, and financial plans for college.

Overall, there was less focus on returning students and more priority placed on first-year/first-time entering students. Two institutions indicated the planned pilot of a College Success course specifically for new students. While it is important to begin retention efforts as soon as students arrive on campus, attention must also be paid to those students already enrolled, who may benefit from a targeted intervention strategy as well.

## Details of the individual institutional intervention strategies

	Targeted population(s) of intervention	Details of intended intervention	How success of the intervention will be measured
Institution A	<ul style="list-style-type: none"> <li>i) all students who are performing poorly in their classes and/or not attending as indicated by Early Alert</li> <li>ii) TN Promise and adult students</li> <li>iii) First-time freshmen, especially students unsure of their major choice</li> </ul>	<ul style="list-style-type: none"> <li>i) Early Alert- Students and their faculty advisors will be notified within the first four to six weeks if they are performing poorly on assignments, have excessive tardiness, and/or excessive absences. Advisors will contact students and offer assistance</li> <li>ii) Academic Maps – TN Promise and adult students will enter the college with a plan to graduation in their first semester of enrollment</li> <li>iii) College Success Course Pilot – First-time freshmen enrolled in the pilot course will be exposed to different careers, become familiar with services available to assist them in meeting goals, and develop good habits for successful students.</li> </ul>	<ul style="list-style-type: none"> <li>i) Retention of students who receive alerts will be a 50% or more</li> <li>ii) TN Promise and adult students with Academic Maps will be retained and stay on course to graduate</li> <li>iii) First-time freshmen, especially students who are unsure of their major choice will decide on a major, be retained, and complete their degree in a timely manner.</li> </ul>
Institution B	<ul style="list-style-type: none"> <li>i) Students who have not graduated</li> <li>ii) Nursing students</li> <li>iii) Those who could change their major and be on track to graduate</li> </ul>	Giving advisors a form that includes some information from the [trajectory analytics] data	By count – how many students graduate from the list
Institution C	<ul style="list-style-type: none"> <li>i) First-time entering students (beginning Fall 2016)</li> <li>ii) Nursing students with test grades below 78</li> </ul>	<ul style="list-style-type: none"> <li>i) Mandatory advisement with a success coach and faculty member to review exams, class attendance, study skills and time management, completion of remediation assignments in area of particular weakness; with follow-up monitoring by faculty member and/or dean of Nursing</li> <li>ii) Training for newly hired success coaches who will work with new students to develop academic plan, career plans, and financial plans</li> </ul>	<ul style="list-style-type: none"> <li>i) Improved test scores, improved NCLEX performance</li> <li>ii) Improved retention/graduation of cohort against baseline prior to mandatory advisement of first-time students</li> </ul>
Institution D	Any student previously reported on Faculty Feedback forms during Fall 2014 or Spring 2015	Based on information collected, [providing a referral for] tutoring, counseling, individual meetings with counselors, and etc.	Review the trajectory data for students reported through faculty feedback to look for a positive measure
Institution E	250 College of Science and Mathematics students with declared majors of unlikely success	Professional staff led proactive/intrusive redirect advising outreach with subsequent conversations based on motivational interviewing techniques, Prochaska's Stages of Change model, and appreciative advising techniques.	<p>A student moving at least one stage on the Stages of Change spectrum is documented with and forward movement considered an interim success.</p> <p>A student changing his/her major from one of unlikely success to one of likely success is considered and documented as a success.</p>

## Decision-Making Culture

Understanding the culture around decision-making, specifically the use of data in decision-making, is an important first step to building the institutional capacity for utilizing analytics to its fullest potential. Baer and Norris propose a Three-Stage Model for Capacity Building for Analytics, noting that:

- institutions need to take different approaches depending on their stage of development, and
- without concerted action, institutions will fall behind.

The second stage of this research project sought to gauge the existing capacity for data-informed/data-driven decision-making at TBR institutions. Survey questions were sent to all thirteen community colleges and six universities via the chief academic officer. Ten community colleges and three universities responded.

When asked, *“Does your institution have a culture of data-informed decision making?”* each of the participating institutions responded affirmatively. Three institutions indicated growth and increasing usage of data in academic affairs –data had previously been used to make decisions in areas such as enrollment management and Quality Assurance planning. Two of the institutions described a strong commitment to using data in making decisions related to college processes. One school in particular noted the partnership between the offices of Institutional Research and Information Technology in gathering information for making decisions.

## ACCESS TO DATA AND INSTITUTIONAL PREPAREDNESS

Beyond a willingness of key personnel to utilize data in the decision-making process, institutions must also have access to and resources for handling the types of large datasets necessary for analytics. As evidenced in the survey responses, having a culture of data-informed decision-making is not necessarily dependent on the institutions’ access to data or level of preparedness.

In particular, the survey asked, *“Do you feel that you have easy access to the data needed to inform decision-making?”* (Surveys were sent to the chief academic officer, but some schools had the Institutional Research office respond to the questions, based on expertise.) There was markedly less consensus in these responses: only 6 out of 13 respondents indicated yes.

Institutions noted that the data needed exists, but it is not organized in a format that is conducive to analysis. Data is siloed in different databases, within different systems – for example early alert data is separated from Banner information.

The source of data and reports for institutions overwhelming came from Argos. Other sources included automated email reports, dashboards, and specialized generators in the office of Institutional Research.

It is important to note that analytics is not about generating more data – it is mining the existing data for insights and trends that can be turned into action plans and strategies. In the survey, a majority indicated that they have “enough” data at their institution to effectively do analytics.

In response to, *“Is your institution prepared to handle big data and analytics?”* only three schools indicated strongly that resources and personnel were in place at the institutional level. Those schools had purchased analytics software, were training staff members, and had data mining tools and a functional data store for analysis.

However, among the ten other institutions, common themes for the lack of preparedness in their explanations arose.

---

We all do many jobs, so having time to do a careful analysis is difficult, especially with big data.

Several institutions indicated the size of the job of analytics as a particularly daunting obstacle. One school noted that with varying software systems available (Excel, Access, SPSS), it is very much a manual process to pull data from sources such as Banner into workable formats.

Another school stated that although the institution is in the process of experimenting with predictive models, much of their time (in that office) is devoted to state and national data requests and ad hoc reporting. There is an overall perception across the institutions of TBR that undertaking analytics is a time-, and resource-consuming process.

---

More resources (personnel and expertise) are needed in order to take full advantage of big data analytics.

The second most common theme in discussing the preparedness of the institution was that of a lack of human capital necessary to successfully carry out analytics. Not having adequate staff, lack of expertise in data science and advanced statistical methods among the current personnel, and a desire for more training were all mentioned by multiple institutions. Additionally, several schools indicated the hope that the TBR Central Office would be able to offer centralized programming, access to the CDR, and manpower in general for statistical analysis.

## LEVERAGING SYSTEM DATA

The probability of reliably and confidently predicting behaviors and outcomes is improved when using predictive tests and analyses on a single, massive set of records coming from a diverse array of contributors normalized around common data definitions.

Organizations like PAR Framework take advantage of this axiom of data mining and predictive analytics by inviting institutions to provide de-identified student data in exchange for analytics reports. PAR Framework combines data from a variety of institutions – public and private, four year and two year, large and small student populations – to develop sophisticated models of student success that member institutions can take advantage of for their own optimization.

TBR institutions are the position to provide a similar capability, in house. The 19 institutions represent a wide variety of students and have the benefit of already using a common nomenclature: for example most schools have similar definitions of student demographics (non-traditional, low-income/Pell eligible, etc.) and use a common course numbering system.

In the survey, institutions were asked, *“Do you feel that there is an advantage in leveraging system data to inform institution analytics?”* Responses again indicated an overall lack of awareness of the benefits of group aggregation for predictive analytics. Specifically, many schools responded that data from other schools would be of limited use – mainly for transfer students and comparison of benchmarks to other institutions.

Institutions did note other benefits to utilizing the Central Office in data analysis: the manpower of the Central Office for statistical analysis and management of institutional data to decrease replication of work for institutions.

## Conclusions and Recommendations for Practice

---

According to the research of EDUCAUSE, conversations within higher education institutions about data, information, reporting, and analytics tend to begin with a focus on technology. While these elements are certainly necessary for successful implementation, they are not sufficient alone to the ultimate success of using analytics to optimize student success.

Instead, “the truly strategic issue facing higher education today is [...] the ability of individual institutions and the higher education industry as a whole to deeply acquire in a purposeful and continuous manner the full set of organizational capacity and behaviors needed to optimize student success.” (L. Baer and D. Norris, 2013)

TBR institutions interested in harnessing the power that predictive analytics can provide must be willing to transform the culture of decision-making from one of simply reporting with data to

a culture of evidence, where data is able to provide actionable information and insights into student success.

## ASSESSING INSTITUTIONAL MATURITY

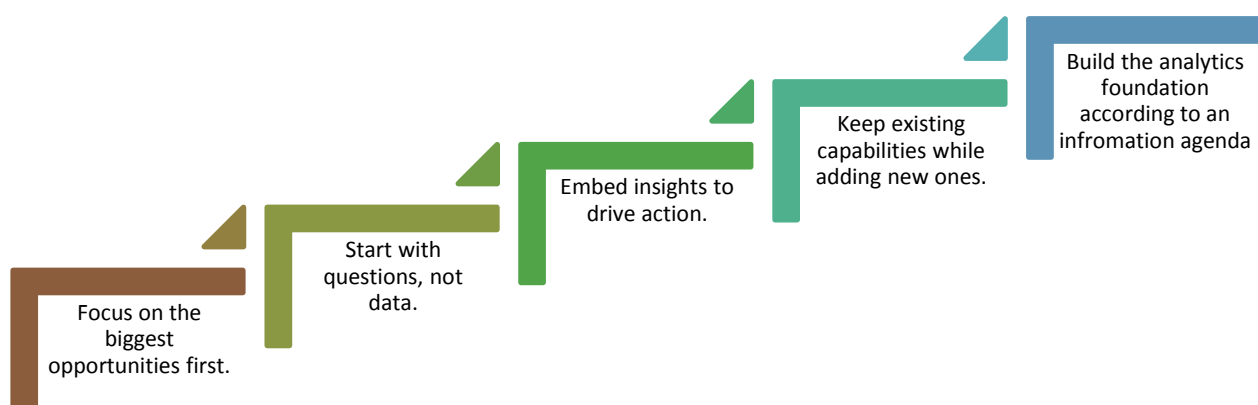
The EDUCAUSE Center for Applied Research (ECAR) has developed an analytics Maturity Index questionnaire that measures in five dimensions the readiness of an institution for effective use of analytics:

1. CULTURE – Committed leadership; culture accepts use of data to make decisions
2. DATA & TOOLS – Clean, standardized data and reports; right tools and software
3. INVESTMENT – Funding and staffing for analytics
4. EXPERTISE – IR and/or business professionals with analytics training
5. INFRASTRUCTURE – Storage capacity; IT professionals supporting analytics; policies regarding security and rights to data

The issues as reported above by TBR institutions point to an immaturity in dimensions 2, 3, and 4. Strengthening these areas would require institutional commitment to standardizing data and reports, investing in software tools and training for that software, viewing the funding of analytics as an investment in the future of the college, and acquiring dedicated professionals with specialized analytics training.

## RAISING ANALYTICS IQ

Today, the “analytics IQ” of rank-and-file leaders in higher education typically lags behind the new and rapidly accelerating developments in the analytics field. As evidenced in the survey responses, institutional leaders tend to overestimate their organization’s capacity in data, information, and analytics, or have mismatching perceptions of their capabilities and resources.



There is substantial need to increase professional development, capacity building, and the

analytics IQ of institutional leadership and practitioners, at all levels across the system. The steps in the graphic above, adapted from a presentation by Strategic Initiatives Inc., outlines a suggested path for strategic planning in institutions looking to increase awareness of analytics among leadership. The importance of including analytics as part of an overall process can be seen here as well: focusing on opportunities and beginning with the important questions first.

The analytics process as outlined by EDUCAUSE involves the following steps:

1. Start with a strategic question
2. Find/collect the appropriate data to answer that question
3. Analyze that data with an eye towards prediction and insight
4. Formulate and present in ways that are understandable and actionable
5. Feedback into the process of addressing strategic questions and identifying new ones

Increasing the analytics IQ will require a cross-institutional partnership of academic and student affairs; institutional research and information technology. There also must be significant buy-in from faculty and other staff such as advisors, counselors, and others involved in student support. This will promote the institutional shift to a culture that incubates and encourages decision-making driven by data and analytics.

Among TBR institutions, successful implementation of strategies developed from the trajectory analytics report can have a ripple-outward effect, encouraging future use of predictive models and individual development of new ones. The current analytics report can also help increase awareness of the benefits of aggregated data from multiple institutions beyond comparison capacity.