Data for Students: The Potential of Data and Analytics for Student Success

Katherine Leu

Key Findings
- Data analytics and data science can address challenges to student success.
- Postsecondary institutions have already creatively used analytics to address the problem of college completion through innovations such as academic early warning systems and adaptive learning technologies.
- A diverse array of data on postsecondary education exists both within and outside of institutions and can be used in analytics to provide a richer view of student success and improve equity.
- Increased data collection and analysis open up the challenges of data linkage across units and the risk of ethical and privacy violations, which deserve more attention.

Data have proliferated in seemingly every area of life. "Big data" and the algorithms that make sense of them have revolutionized fields like medicine and business[1] and have led to the rise of data analytics and data science, which use visualization, algorithms, and other data analysis techniques to extract insights from data and drive decisionmaking.[2] Postsecondary education, too, is experiencing its own data deluge. Online applications now capture fine-grained teaching and learning data on a large scale,[3] and large quantities of institutional data have been collected in response to increased external pressure for accountability in higher education.[4]

Student success is a key area where these data can be put to use. Student success can be defined broadly to include students' personal development and goals; however, efforts to foster student success have generally focused on students' academic performance.[5] College completion in particular has gained significant attention since it was made a focus of the Obama administration's higher education policy. However, although completion rates have improved, concerns about equity have not been adequately addressed.[6] Other issues also affect students' success. Costs of college and debt burdens are both rising,[7] as is anxiety about the value of a postsecondary credential in the labor market.[8] And even with the abundance of available data, providing relevant, timely information to the students and families who need it remains a challenge.[9] With their ability to deliver insights quickly from large troves

of data, data analytics and data science have the potential to address some of the most pressing challenges to student success.

In this brief, I discuss the potential of new data sources in postsecondary education for data analytics and data science focused on student success, with the aim of sharing this knowledge with stakeholders in postsecondary education and with data scientists. Postsecondary stakeholders could benefit from understanding the new possibilities that data science may provide for addressing student success issues, and data scientists could benefit from knowing about the opportunities and analytic questions in postsecondary education. I first describe how institutions have applied data analytics to student academic data to address problems related to completion. I then discuss the range of data that exist both within and outside institutions and provide a taxonomy of how these data may be used. I argue that incorporating a broader array of data into analytics could provide richer answers to questions about how to improve student success. Finally, I conclude with a discussion of the practical and ethical issues that accompany increased data collection and analysis.

Academic Analytics for College Completion

In recent years, more and more postsecondary institutions have begun using analytics to aid in institutional operations; this growing field is termed "academic analytics."[10] Although the goals of academic analytics have been diverse, serving postsecondary institutions' business and accountability goals in admissions, fundraising, and student affairs,[11] a major area of application has been in student success, specifically college completion. I discuss examples of these applications subsequently.

Georgia State University has been recognized as a leading institution on this front. Using predictive analytics, among other innovations, it raised its 6-year graduation rate from 32 percent in 2003 to 55 percent in 2018 and reduced graduation rate gaps between disadvantaged and non-disadvantaged students.[12] At the core of this strategy is a predictive analytics system tracking more than 800 risk factors of dropout.[12,13] A team of advisers monitors this system to provide targeted help to students.[12,13] Informed by insights from this predictive analytics system, the university has also clarified student academic pathways and made changes to the structure of course requirements.

Other institutions also have developed tools to facilitate students' academic progress. At Austin Peay State University in Tennessee, a tool called Degree Compass provides "Netflix-style" course recommendations for students that account for their interests, program requirements, and past performance to help them tackle the challenge of choosing courses from a wide array of options.[14] Tools like Course Signals at Purdue and ECoach at the University of Michigan help students manage their performance within courses by providing personalized advice and early warnings of poor performance.[15,16] Universities have often partnered with external vendors to build these tools. The "Check My Activity" tool at the University of Maryland, Baltimore County, was developed in collaboration with Blackboard, a course management system, and allows students to compare their level of activity in Blackboard against an anonymous summary of their course peers' performance.[17]

Alongside work in academic analytics, work in the field of learning analytics has focused on using data-driven tools and models to analyze online learning behavior and provide personalized solutions to students' learning challenges.[18] This work has led to the development of adaptive learning technologies, which use machine learning to adapt online learning experiences in real time to each student's individual skills and understanding of the material.[19] Institutions have begun to implement these technologies: for example, the University of Central Florida and Colorado Technical University partnered with Realizeit, an online learning vendor, to pilot an adaptive learning platform in select courses. Early evidence shows improvement in student outcomes, such as greater engagement with course concepts.[20]

Alternative Data Sources and Their Applications for Student Success

The above examples demonstrate that higher education institutions have made strides in harnessing student academic data and classroom and learning data to promote student success and completion. However, barriers to completion are often nonacademic: many students face challenges such as food and housing insecurity,[21] as well as employment and family demands that compete for their time and hinder their ability to obtain a credential.[22] Linking students' nonacademic records to their academic records could provide insight into the ways these challenges systematically affect academic progress, and could also be used as early warning indicators. Institutions could also evaluate the effectiveness of the student services they offer using linked student services and academic data.

Digital tools also could offer solutions directly to students; for example, such tools could connect students to available housing or sources of food, or help students navigate services the institution already provides. Many schools offer emergency aid, for instance, but students are often unaware of such programs.[23] Researchers have begun to make creative use of new data sources within institutions to study aspects of students' experiences outside of class that can affect

completion. Studies of the digital traces of campus social life, such as key card swipes into campus buildings, have examined their potential to predict student engagement and integration, which are ultimately crucial to retention and completion.[24]

To make sense of the growing diversity of data sources, Table 1 displays the types of data sources available within postsecondary institutions, as well as student success issues that can be addressed by each data type. *Classroom and learning data* pertain to higher education's core function of educating students and measuring learning. *Student academic records* measure students' progress through their academic programs. *Student nonacademic records* pertain to students themselves and to the nonacademic aspects of students' lives. *Institutional unit data* relate to the operations of an institution's academic programs and business units.

In addition, a vast amount of postsecondary education data are being collected outside of individual institutions, and these data, and the tools and analyses that make sense of them, can also be brought to bear on student success. Diverse sources of data can provide a broader, more holistic understanding of what student success entails and the factors that shape it: student success could encompass not only the completion of a degree or certificate program, but also skills or knowledge acquisition, finding a job, and socioeconomic mobility.

Data linking academic experiences to workforce and economic data can be brought to bear on these issues. State governments have begun to build State Longitudinal Data Systems that link unit-level K–12, postsecondary, and workforce data, as well as tools and dashboards to explore these troves of data.[8] Information from these systems connecting students' college experiences to their labor market outcomes would be invaluable in informing job-seeking students, as well as researchers and policymakers, about which courses, majors, and other experiences can lead to certain occupations or industries. Meanwhile, the economic aspects of postsecondary education have well-documented effects on student outcomes,[7] but despite the abundance of data displayed in online net price calculators[25] and college search websites for prospective students,[26,27] students still suffer from a lack of clear and targeted information that would enable them to navigate this complex financial landscape.[9] These comprehensive data on costs, tuition, and financial aid could be made more interpretable and personalized, and could also be integrated into institutional systems to better inform students of the personal financial implications of their degree plans.

Table 2 below displays the postsecondary data topics, data sources, and student success issues that can be addressed by data sources across and outside of postsecondary institutions. *Institution-level data* provide information about a given

Table 1. Data within postsecondary institutions and student success issues they can address

Data Type	Data Sources	Student Success Issues
Classroom and learning data	Online learning platforms, learning management systems, instructors' records	Student learning; curricular improvement; classroom environment and peer effects
Student academic records	Institutions, transcripts, National Student Clearinghouse	Courses that are barriers to completion; academic pathways; personalized advising
Student nonacademic records	Institutions, financial aid organizations	Student social life and integration; connection to resources for basic needs; usage and effectiveness of student services
Institutional unit data	Academic programs and business units within institutions	Composition of student body within units; differences in student outcomes across units

Table 2. Data across and outside of postsecondary institutions and student success issues they can address

Data Type	Data Sources	Student Success Issues
Institution-level data	College Scorecard, Integrated Postsecondary Education Data System (IPEDS)	Differences between institutions in graduation and retention rates; differences between institutions in tuition, financial aid, and debt
Data on linkages to K–12 and to workforce	State Longitudinal Data Systems, employers	High school preparation for college success; student employment and earnings; credentials or experiences related to careers
Governance data	Governments, professional associations (e.g., State Higher Education Executive Officers Association [SHEEO])	Budget appropriations for educational institutions; laws and policies across education systems

postsecondary institution and its performance relative to other institutions. *Data on linkages to K–12 and to workforce* provide information on students' transitions between education levels and the workforce, and how postsecondary education relates to K–12 education and workforce needs. *Governance data* pertain to the activities of leaders and governments that affect postsecondary education.

Finally, higher education must continue to address the issue of equity, both within and between postsecondary institutions. Inequalities in student outcomes by race and class persist, as do resource differentials between institutions,[6] and institution-level data and institutional unit data on student outcomes can illuminate the patterns and causes of differences at higher levels of analysis. Institutional Mobility Report Cards, for instance, use tax records linked to institution-level data from the Department of Education's College Scorecard to characterize an institution's contribution to promoting intergenerational income mobility for its students.[28] Governance data, such as data on federal funding for higher education institutions, have received relatively little attention in their implications for inequality, but deserve more focus.[7] Along similar lines, work also should focus on ensuring that institutions with fewer resources can still implement and reap the benefits of robust data analytics for their students.

Challenges

"Big data," analytics, and data science show immense promise for working toward student success in postsecondary education. Work should continue to develop and expand the application of these approaches to understanding and addressing both academic and nonacademic issues in students' lives, as well as to alleviating inequities. However, large-scale data collection, linkage, and analysis also open practical and ethical challenges.

One important challenge lies in building the capacity and infrastructure to create data linkages across institutional units or organizations. Data necessary for an analysis may be scattered across separate entities that do not communicate with each other, contributing to disorganized, inconsistent, or incomplete data.[29] Higher education institutions should continue to build and expand the infrastructure and capacity necessary to collect and analyze data, working within information systems and institutional cultures to align data collection standards, build knowledge, and foster the willingness to adopt new methods.[30] Toward this end, the Gates Foundation and the Institute for Higher Education Policy (IHEP) have spearheaded an effort to map and standardize data reporting on college students' entire postsecondary trajectory.[31] Along similar lines, a closer

dialogue across sectors, for example between higher education administrators, researchers, policymakers, and data scientists, could develop a shared understanding that would encourage progress and innovation in using data to help address the most pressing problems hindering student success.

Large-scale data collection and analysis also raise the issues of data privacy and ethical data collection and use. Even as institutions collect reams of data on their students, students themselves often are unaware of this fact and have not been given the opportunity to provide informed consent or opt out.[32,33] It is imperative that administrators justify the data they collect by weighing the potential benefits to students against the potential harms, actively communicate these considerations to students, and allow students some real measure of control over what happens to their information.[32] The risks of large-scale student data collection can be substantial: as more and more data are collected on individuals, individuals are increasingly vulnerable to the negative consequences of data breaches caused by improper data collection, storage, and analysis practices. These effects are magnified when multiple data sources are linked, potentially making individuals easier to identify.[11] In addition, biased algorithms can unfairly profile individuals on the basis of their characteristics.[11] While work has called for more serious consideration of the legal and ethical implications of large-scale student data collection and linkages,[11,32,33,34] more institutions and stakeholders need to take heed and develop guidelines and policies to ensure transparency, consent, and fairness. Above all, stakeholders in postsecondary education must not lose sight of ethical considerations as they explore the potential of data analytics for improving student success.

References

1. Mayer-Schönberger V, Cukier K. Big data: a revolution that will transform how we live, work, and think. New York (NY): Houghton Mifflin Harcourt; 2013.

2. Loukides M. What is data science? O'Reilly Media, Inc.; 2010. Available from: https://www.oreilly.com/radar/what-is-data-science

3. Picciano AG. The evolution of big data and learning analytics in American higher education. J Async Learn Network. 2012;16(3):9–20. https://doi.org/10.24059/olj.v16i3.267

4. Campbell JP, DeBlois PB, Oblinger DG. Academic analytics: a new tool for a new era. EDUCAUSE Rev 2007;42(4):40–57.

5. Defining Student Success Data Initiative. Defining student success data: recommendations for changing the conversation. Chicago (IL): Higher Learning Commission; 2018 Dec.

4

6. Teranishi RT, Bezbatchenko AW. A critical examination of the college completion agenda: advancing equity in higher education. In: Martínez-Alemán AM, Pusser B, Bensimon EM, editors. Critical approaches to the study of higher education: a practical introduction. Baltimore (MD): Johns Hopkins University Press; 2015.

7. Goldrick-Rab S. Paying the price: college costs, financial aid, and the betrayal of the American dream. Chicago (IL): The University of Chicago Press; 2016. https://doi.org/10.7208/chicago/9780226404486.001.0001

8. Carnevale AP, Garcia TI, Gulish A. Career pathways: five ways to connect college and careers. Washington (DC): Georgetown University Center on Education and the Workforce; 2017.

9. Bleemer Z, Zafar B. Information heterogeneity and intended college enrollment. Soc Sci Res Net. 2014. Federal Reserve Board of New York Staff Report. https://doi.org/10.2139/ssrn.2477860

10. Baepler P, Murdoch CJ. Academic analytics and data mining in higher education. Int J Schol Teaching Learning 2010;4(2). https://doi.org/10.20429/ijsotl.2010.040217

11. Ekowo M, Palmer I. The promise and peril of predictive analytics in higher education: a landscape analysis. Washington (DC): New America; 2016 Oct.

12. Renick T. Complete college Georgia 2018 status report: Georgia State University. Atlanta (GA): Georgia State University; 2018 Sep.

13. McMurtrie B. Georgia State U. made its graduation rate jump. How? Chron High Educ. 2018 May. Available from: https://www.chronicle.com/article/Georgia-State-U-Made-Its/243514

14. Denley T. Case Study 3. Austin Peay State University: Degree Compass. In: Oblinger DG, editor. Game changers: education and information technologies. Louisville (CO): EDUCAUSE; 2012. pp. 263–7.

15. Arnold KE. Signals: applying academic analytics. EDUCAUSE Quarterly 2010; 33(1).

16. Calvert Mason K. Can just-in-time advice keep more college students on track? PBS NewsHour. 2015 Aug 21.

17. Fritz J. Classroom walls that talk: using online course activity data of successful students to raise self-awareness of underperforming peers. Internet Higher Educ 2011;14(2):89–97. https://doi.org/10.1016/j.iheduc.2010.07.007

18. Ferguson R. Learning analytics: drivers, developments and challenges. Int J Tech Enhanced Learning 2012;4(5/6):304–17. https://doi.org/10.1504/IJTEL.2012.051816

19. Dziuban C, Howlin C, Moskal P, Johnson C, Parker L, Campbell M. Adaptive learning: a stabilizing influence across disciplines and universities. Online Learning J 2018;22(3):7–39. https://doi.org/10.24059/olj.v22i3.1465

20. Dziuban C, Howlin C, Johnson C, Moskal P. An adaptive learning partnership. EDUCAUSE Review; 2017 Dec 18.

21. Goldrick-Rab S, Baker-Smith C, Coca V, Looker E, Williams T. College and university basic needs insecurity: a national #RealCollege survey report. The Hope Center. 2019 Apr. Available from: https://hope4college.com/wp-content/uploads/2019/04/HOPE_realcollege_National_report_digital.pdf

22. Wilbur TG, Roscigno VJ. First-generation disadvantage and college enrollment/completion. Socius 2016;2(1).

23. Kruger K, Parnell A, Wesaw A. Landscape analysis of emergency aid programs. Washington (DC): National Association of Student Personnel Administrators; 2016.

24. Bowman NA, Jarratt L, Polgreen L, Kruckeberg T, Segre A. Early identification of students' social networks: predicting college retention and graduation via campus dining. J Coll Student Dev 2019;60(5):617–22. https://doi.org/10.1353/csd.2019.0052

25. Higher Education Opportunity Act of 2008, Pub. L. No. 110–315, 122 Stat. 3078 (2008 Aug 14).

26. O'Shaughnessy L. The 10 best college websites. Washington (DC): US News and World Report. 2010 Jun 15. [cited 2019 Sep 2]. Available from: https://www.usnews.com/education/blogs/the-college-solution/2010/06/15/the-10-best-college-websites

27. Schonberg J. Start your college search with Google. Mountain View (CA): Google LLC. 2018 Jun 12. [cited 2019 Sep 2].Available from: https://www.blog.google/products/search/college-search-google/.

28. Chetty R, Friedman JN, Saez E, Turner N, Yagan D. Mobility Report Cards: the role of colleges in intergenerational mobility. National Bureau of Economic Research. 2017 Dec. NBER Working Paper. https://doi.org/10.3386/w23618

29. Daniel B. Big data and analytics in higher education: opportunities and challenges. Br J Educ Technol 2015;46(5):904–20. https://doi.org/10.1111/bjet.12230

30. Cardoza D, Gold J. Moving from data to action: changing institutional culture and behavior. In: Gagliardi JS, Parnell A, Carpenter-Hubin J, editors. The analytics revolution in higher education: big data, organizational learning, and student success. Sterling (VA): Stylus Publishing; 2018. pp. 119–38.

31. Engle J. Answering the call: institutions and states lead the way toward better measures of postsecondary performance. Seattle (WA): Bill and Melinda Gates Foundation; 2016.

32. Rubel A, Jones KML. Student privacy in learning analytics: an information ethics perspective. Information Society 2016;32(2):143–59.

33. Gilliard C. How ed tech is exploiting students. Chron Higher Educ 2018 Apr. Available from: https://www.chronicle.com/article/How-Ed-Tech-Is-Exploiting/243020

34. Sun JC. Legal issues associated with big data in higher education: ethical considerations and cautionary tales. In: Lane JE, editor. Building a smarter university: big data, innovation, and analytics. Albany (NY): SUNY Press; 2014. pp. 27–55.

5

About the Author

Katherine Leu, BA, is a doctoral student in the Department of Sociology at the University of Michigan in Ann Arbor. She was formerly an education analyst in the Center for Postsecondary Education at RTI International.

RTI Press Research Briefs and Policy Briefs are scholarly essays on policy, methods, or other topics relevant to RTI areas of research or technical focus.

RTI International, 3040 East Cornwallis Road, PO Box 12194
Research Triangle Park, NC 27709-2194 USA

+1.919.541.6000   rtipress@rti.org   www.rti.org

www.rti.org/rtipress