Data Analytics in Higher Education: Key Concerns and Open Questions

Alan Rubel

Kyle Jones

DATA ANALYTICS IN HIGHER EDUCATION: KEY CONCERNS AND OPEN QUESTIONS

ALAN RUBEL* AND KYLE JONES**

"Big Data" and data analytics affect all of us.1 Data collection, analysis, and use on a large scale is an important and growing part of commerce, governance, communication, law enforcement, security, finance, medicine, and research. And the theme of this symposium, "Individual and Informational Privacy in the Age of Big Data," is expansive; we could have long and fruitful discussions about practices, laws, and concerns in any of these domains. But a big part of the audience for this symposium is students and faculty in higher education institutions (HEIs), and the subject of this paper is data analytics in our own backyards. Higher education learning analytics (LA) is something that most of us involved in this symposium are familiar with. Students have encountered LA in their courses, in their interactions with their law school or with their undergraduate institutions. Instructors use systems that collect information about their students, and administrators use information to help understand and steer their institutions. More importantly though, data analytics in higher education is something that those of us participating in the symposium can actually control. Students can put pressure on administrators, and faculty often participate in university governance. Moreover, the systems in place in HEIs are more easily comprehensible to many of us because we work with them on a day-to-day basis. Students use systems as part of their course work, in their residences, in their libraries, and elsewhere. Faculty deploy course management systems (CMS) such as Desire2Learn, Moodle, Blackboard, and Canvas to structure their courses, and administrators use information gleaned from analytics systems to make operational decisions. If we (the participants in the symposium) indeed care about Individual and Informational Privacy in the Age of Big Data, the topic of this paper is a pretty good place to hone our thinking and put into practice our ideas.

This paper is organized as follows. In part I, we provide an overview of data analytics in higher education, noting some ways in which information is gathered, analyzed, and used. In part II, we address a number of concerns based on the account provided. And in part III, we outline a series of next research steps necessary for moral and legal analysis of higher education analytics.

I.   UNDERSTAND HIGHER EDUCATION ANALYTICS

The idea of data analytics is becoming familiar, and use of data analytics in higher education is becoming a matter of increasing public interest. A recent report by the New America Foundation about the state of higher education analytics frames the subject by contrasting two recent cases. [2] The first case concerns how Georgia State University (GSU) examined millions of student records over the course of a decade in order to create a list of "red flag" conditions that are correlated with a decreased likelihood of a student graduating. Among these are: receiving a poor grade in a course needed for one's major, failure to take a required course at the appropriate time, and registering for a course not relevant to one's major. When such red flags turn up, it prompts an in-person meeting between the student and his or her advisor. Because the system was geared toward advising interventions, GSU also made a substantial investment by hiring many new academic advisors and thereby dramatically decreasing the advising load of each advisor. This combination of analytics and greater investment in advisors to make use of those analytics is correlated with increased graduation rates, shorter times to degree, and "[l]ow-income, first-generation and minority students have closed the graduation rate gap."[3]

So far, so good!

A different, recent national news story reflects a different approach to analytics in higher education. In early 2016, the president of Mount St. Mary's University received widespread attention for a plan to increase the college's retention and graduation rates by using analytics, including a personality survey, to identify students likely to drop out, and encouraging them to leave early (i.e., before they would count in the retention measures).[4] The president at the time, Simon Newman, resigned shortly

---

[2] Manuela Ekowow & Iris Palmer, *The Promise and Peril of Predictive Analytics in Higher Education: A Landscape Analysis*, NEW AMERICA (Oct. 24, 2016), https://na-production.s3.amazonaws.com/documents/Promise-and-Peril_4.pdf;
Anya Kamenetz, *How One University Used Big Data to Boost Graduation Rates*, NATIONALPUBLICRADIO: NPRED   (Oct.   30,   2016,   5:03   AM), http://www.npr.org/sections/ed/2016/10/30/499200614/how-one-university-used-big-data-to-boost-graduation-rates.

[3] Kamenetz, *supra* note 2.

[4] Susan Svrluga, *University President Allegedly Says Struggling Freshmen are Bunnies That Should   be   Drowned,*   WASHINGTON   POST   (Jan.   19,   2016),

after this incident became public. The main outcry appeared to be his comments that the university needed to stop thinking of first-year students as "cuddly bunnies," and that "[y]ou just have to drown the bunnies … put a Glock to their heads."[5] The point of using the example here, though, is not the abhorrent metaphor or callous indifference to students' struggles, but the attempt to use analytics to achieve better statistics. That is, Newman's goal was to have better retention rates without addressing the underlying issue of student experience.

These two different stories reflect an underlying tension. There are obvious potential benefits to expansive use of data analytics in higher education (e.g., the Georgia State case). However, the allure of doing better along some metrics may mask ways in which data may be used to undermine other values. To understand that, it will be useful to describe in a little more detail analytics practices in HEIs.

What is learning analytics?

Learning analytics is a socio-technical data-mining practice primarily employed in educational contexts. It is technical in the sense that information technologies are used to collect information, but it is a social practice in that where and how information collection occurs, how information is analyzed, and how conclusions are deployed depend on social practices. LA is prevalent at all levels of education, but this paper leaves to one side use of analytics in K-12 education as the policy pressures there are quite different. A precise definition of learning analytics has been hard to pin down.[6] This is due in part because learning analytics includes a "sprawling" array of information and practices, the ways in which learning analytics practices have evolved with information technology, and the fact that goals of learning analytics change as data-driven initiatives have

---

https://www.washingtonpost.com/news/grade-point/wp/2016/01/19/university-president-allegedly-says-struggling-freshmen-are-bunnies-that-should-be-drowned-that-a-glock-should-be-put-to-their-heads/; Rebecca Schisler & Ryan Golden, *Mount President's Attempt to Improve Retention Rate Included Seeking Dismissal of 20-25 First-Year Students,* THE MOUNTAIN ECHO (Jan. 19, 2016), http://msmecho.com/2016/01/19/mount-presidents-attempt-to-improve-retention-rate-included-seeking-dismissal-of-20-25-first-year-students/.

[5] Ekowow & Palmer, *supra* note 2; Svrluga, *supra* note 4; Schisler & Golden, *supra* note 4; Scott Jasnchik, *The Questions Developed to Cull Students,* INSIDE HIGHER ED (Feb. 12, 2016), https://www.insidehighered.com/news/2016/02/12/questions-raised-about-survey-mount-st-marys-gave-freshmen-identify-possible-risk.

[6] John P. Campbell et al., *Academic Analytics: A New Tool for a New Era.* 40 EDUCAUSE REVIEW 41, July–Aug. 2007, http://net.educause.edu/ir/library/pdf/erm0742.pdf; Adam Cooper, *What is Analytics? Definition and Essential Characteristics* 1 CETIS ANALYTICS SERIES 5, http://publications.cetis.ac.uk/2012/521; Angela van Barneveld et al., *Analytics in Higher Education: Establishing a Common Language*, 47 EDUCAUSE REVIEW 1, Jan.–Feb. 2012, http://net.educause.edu/ir/library/pdf/ELI3026.pdf.

opened up new opportunities. 7 The lack of a single definition is not surprising. Davenport argues that terms like 'decision support,' 'executive support,' 'business intelligence,' and 'Big Data' all generally refer to how organizations and institutions seek to turn data into actionable information via advanced computing practices.8 The differences between the terms is largely a matter of relevant social values, affordances of technological systems, and the types of political and economic gains that actors seek. For the purposes of this paper, Siemens' definition will suffice: "Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs."9

What are the purposes of learning analytics?

Advocates point to two primary ends that learning analytics furthers. First, they understand LA as a form of institutional business intelligence that draws on the financial, operational, and academic needs and interests found in educational contexts. Hence, advocates argue that educational institutions can use LA to inform institutional practices in order to optimize resource use and find new sources of income for universities in a time of austerity.10 They also understand LA as a political tool that can be useful in making data-driven arguments in the face of growing pressures for education "accountability."11 Advocates hope to use analytics to counter criticisms that American higher education has lost ground in international rankings,12 that HEIs have not ensured that members of under-represented groups achieve learning outcomes at satisfactory rates,13 and that higher education is not meeting market demands for highly-skilled employees (i.e., they are not closing the purported "skills gap").14 The hope is that education

---

7 GEORGE SIEMENS, *Learning Analytics: Envisioning a Research Discipline and a Domain of Practice*, PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 4-8 (Simon B. Shum et al. eds., 2012).

8 Thomas H. Davenport, BIG DATA AT WORK: DISPELLING THE MYTHS, UNCOVERING THE OPPORTUNITIES 10 (Harv. Bus. Pub. Corp. 2014).

9 Siemens, *supra* note 7.

10 Philip J. Goldstein & Richard N. Katz, *Academic Analytics: The Uses of Management Information and Technology in Higher Education,* 8 EDUCAUSE CENTER FOR APPLIED RESEARCH (2005), https://net.educause.edu/ir/library/pdf/ers0508/rs/ers0508w.pdf.

11 Van Barneveld *supra* note 6; DARRELL M. WEST, BIG DATA FOR EDUCATION: DATA MINING, DATA ANALYTICS, AND WEB DASHBOARDS (Christine Jacobs & Stephanie Dahle eds., 2012), https://www.brookings.edu/wp-content/uploads/2016/06/04-education-technology-west.pdf.

12 Ellie Bothwell, *US Dominance in the World University Rankings 2015-16*, TIMES HIGHER EDUC., (Sept. 30, 2015), https://www.timeshighereducation.com/news/big-beasts-strive-to-thrive-in-shifting-environment; OECD, EDUCATION AT A GLANCE 2014 OECD INDICATORS (17th rev. ed. 2014), http://www.oecd-ilibrary.org/education/education-at-a-glance-2014_eag-2014-en.

13 TERRIS ROSS ET. AL., U.S. DEP'T OF EDUC., HIGHER EDUCATION: GAPS IN ACCESS AND PERSISTENCE STUDY (Nat'l Ctr. For Educ. Stat. 2012), https://nces.ed.gov/pubs2012/2012046.pdf.

14 James Bessen, *Employers Aren't Just Whining – the "Skills Gap" is Real,* THE HARV. BUS.

data and analytics will "address [these] intractable challenges" in higher education.[15]

A second aim of LA is to better understand learner performance, both in terms of current progress and predicted success.[16] To do so, LA systems analyze data and information about students in order to "tailor educational opportunities" to their needs. [17] Sometimes, personalized learning experiences are created via algorithms and presented in applications; other times, LA informs instructors and other institutional actors of student needs in order to develop customized interventions. All of this is made possible by aggregating and analyzing troves of student data to "make visible" student behaviors once "unseen, unnoticed, and therefore unactionable."[18] The hope is that in doing so, LA will enable institutional actors to interrogate relationships and patterns related to learning.

Drawing on our working definition of learning analytics above, we can get a better sense of how the two aims of LA overlap. While LA is "first and foremost concerned with learning" it is also about improving educational institutions.[19] The data aggregated for analytic purposes serves to better understand students and their learning behaviors, but it also acts as a means by which a wide range of institutional actors can create actionable information to improve their own practices and programs.[20] Within these two broad aims, there is a range of goals. Where learning is concerned, proponents maintain that learning analytics will:

- help instructors and advisors monitor and analyze student progress;
- predict a learner's future performance, and provide a basis for strategic intervention strategies;
- allow for targeted tutoring strategies;
- develop highly specific feedback for assessment purposes;
- personalize learning with just-in-time lessons;
- and imprint on students the value to be gained by analyzing and reflecting on one's data in order to promote self-

REV., (Aug. 25, 2014), https://hbr.org/2014/08/employers-arent-just-whining-the-skills-gap-is-real/; John F. Ebersole, *Mind the Gap… Between Grad Skills and Employer Expectations*, THE NEW ENGLAND J. OF HIGHER EDUC., (Dec. 2, 2014), http://www.nebhe.org/thejournal/mind-the-gap-between-grad-skills-and-employer-expectations/.

15 Campbell, *supra* note 6, at 42.

16 Johnson L. Smith, et al., The New Media Consortium, The 2011 Horizon Report (2011).

17 *Id*. at 28.

18 Office of Educ. Techn., No. ED-04-CO-0040, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, IX (2012).

19 Doug Clow, *An Overview of Learning Analytics,* 18 Teach. High. Educ. 683, 687 (2013).

20 Mohamed Amine Chatti, et. al., *Learning Analytics: Challenges and Future Research Directions*, 10 ELEED (2014), https://eleed.campussource.de/archive/10/4035.

regulation of behaviors and habits.[21]

With regard to institutional goals, Long and Siemens argue that LA may:

- improve institutional decision making as a means to the end of using resources more efficiently and effectively;
- similarly, enhanced data flows will advance the speed at which decisions are made, which can boost productivity;
- increase institutional transparency by creating datasets and information summaries of those datasets for consumption by stakeholders and the public;
- and may help to better assess the value faculty add to the institution by providing the data necessary to analyze teaching quality, individual reputation, and the reach and impact of disseminated research.[22]

Examples and pervasiveness of data analytics in higher education.

Higher education institutions have developed a wide array of analytic practices based on student data. That's not surprising given the wide range of goals advanced by advocates and the degree of confidence that advocates have in data mining as a tool for educational research and solution to institutional concerns. As a result, institutions use LA in a variety of ways.

Pressures for admissions officers to meet target enrollment numbers have led some HEIs to consider ways that Big Data can reap bigger freshman classes and bring "clarity to a cloudy crystal ball." [23] According to Rivard's reporting on the use of data analytics in admissions, recruiters are stratifying and analyzing personal information of potential applicants in order to "target them for certain traits," including income and ethnicity.[24] Further, admissions professionals purchase and analyze datasets sold by the National Research Center for College and University Admissions, the College Board, or by ACT in order to develop predictive algorithms to score whether or not a specific student is likely to enroll given her profile information. In aggregate, the datasets include millions of student names and identifiable information; it is worth about 37 cents per name.[25]

---

[21] *Id.*

[22] Phil Long & George Siemens, *Penetrating the Fog: Analytics in Learning and Education*, EDUCAUSE REV. 5, Sept.-Oct. 2011, https://net.educause.edu/ir/library/pdf/ERM1151.pdf.

[23] W. Kent Barnds, *Does Big Data Know Best? NSA and College Admissions*, THE HUFFINGTON POST (June 19, 2013), http://www.huffingtonpost.com/w-kent-barnds/does-big-data-know-best-n_b_3460096.html

[24] Ry Rivard, *Political Campaign-Style Targeting Comes to Student Search*, INSIDE HIGHER ED (Oct. 24, 2013), https://www.insidehighered.com/news/2013/10/24/political-campaign-style-targeting-comes-student-search.

[25] Ry Rivard, *Colleges Now Often Rely on Data, Rather Than Gut, in Hunt for Students*, INSIDE HIGHER ED (Sep. 19, 2014), https://www.insidehighered.com/news/2014/09/19/colleges-now-often-rely-data-rather-gut-hunt-students.

In some contexts, analyzing student data has proved more insightful and useful than relying on traditional human labor. Houston Baptist University discovered that its models successfully predicted which students would enroll, regardless of whether or not they received costly viewbooks and mailers–the traditional way of marketing to students.[26] Wichita State University used admissions analytics to forego hiring admissions consultants. Their analytics predicted "high-yield" (i.e., likely to enroll) better than their consultants (96% success rate for the algorithm, 82% for the consultants).[27] Augustana College tracks every communication with a potential student, including via e-mail, Twitter, and Facebook, scoring their messages to rate their "demonstrated interest" in the institution.[28] Other institutions, such as the University of Iowa, invest in Capture Higher Ed's software, which creates profiles full of anonymous usage data about individuals when they visit institutional websites. The software records the IP address of the anonymous user, but attaches it to an identifiable person as soon as that individual clicks on the institution's website from customized e-mails or submits forms with identifiable information. In other words, the previously anonymous profile gets connected with identifiable information and the system gives the individual an engagement score that admissions offices can analyze.[29]

While it is important to the financial success of an institution to get students to campus, it is just as important to retain them through to graduation. So, once students have enrolled, the data mining continues in specific types of applications and systems to keep them engaged and learning. The University of Kentucky sought to increase retention numbers by engaging students with a mobile application. The University's team of data scientists and institutional researchers scoured the data created by students as they interacted with the application. Information collected included, among other things, trails of digital behaviors, ID card swipes at extracurricular events, and sort surveys.[30] The project determined that the

---

[26] *Id.*

[27] Neal Ungerleider, *Colleges are Using Big Data to Predict Which Students Will Do Well--Before They Accept Them* Co.Exist (2013), https://www.fastcoexist.com/3019859/futurist-forum/colleges-are-using-big-data-to-predict-which-students-will-do-well-before-the.

[28] Barnds, *supra* note 23.

[29] Eric Hoover, *Getting Inside the Mind of an Applicant,* The Chron. of Higher Educ. (Sept. 28, 2015), http://www.chronicle.com/article/Getting-Inside-the-Mind-of-an/233403; Amy X. Wang, *Colleges are Spying on Prospective Students by Quietly Tracking Them Across the Internet,* Quartz (Oct. 26, 2015), http://qz.com/513622/colleges-are-spying-on-prospective-students-by-quietly-tracking-them-across-the-internet/.

[30] Carl Straumsheim, University of Kentucky Hopes to Boost Student Retention with Prescriptive Analytics, Inside Higher Ed (Oct. 18, 2013), https://www.insidehighered.com/news/2013/10/18/u-kentucky-hopes-boost-student-retention-prescriptive-analytics.

freshmen-to-sophomore retention rate increased by over one percent. Purdue University is well known for being a leader in LA with their Course Signals system. Course Signals analyzes digital breadcrumbs students leave in the institution's learning management system to predict at-risk rates for poor academic performance and match students with tailored resources. Institutional research on early cohorts who used the system found increased retention and graduation rates, in addition to improved grades.[31]

Data mining has also affected the relationship between students and advisors. Now, advisors can base their guidance to individual students based on insights gleaned from data. E-advising systems increasingly analyze a student's profile in comparison with her peers in order to evaluate her current and predicted rate of success in particular academic programs. These systems then make available to instructors and advisors the student's "electronic reputation" and academic history via dashboards for analysis.[32] Some e-advising tools, like those employed by Austin Peay State University and Arizona State University, create "personalized degree paths," prioritizing automated course recommendations to students based on their academic and professional goals, courses they need to graduate on time, and courses in which students are predicted to be academically successful; these types of systems also have affordances that deny students access to particular courses unless they take specific actions or get advisor approval when their predicted success rate in a course is below a certain threshold.[33]

Colleges and universities are also building capacity to aggregate and analyze data from sensors embedded in and outside campus buildings or, in some cases, on the students themselves. Advocates of LA argue that this type of analysis may be more revelatory of student lives, interactions, and behaviors in learning environments, which cannot be captured in traditional learning management system-based LA systems. [34]   To these ends, institutions are looking into how in-class cameras matched with facial recognition algorithms can automatically take attendance and analyze

---

[31] Press Release, Purdue Univ., Signals Tells Students How They're Doing Even Before the Test (Sept. 1, 2009), http://www.purdue.edu/uns/x/2009b/090827ArnoldSignals.html.

[32] Jeffrey Alan Johnson, The Ethics of Big Data in Higher Education, 21 INT'L REV. OF INFO. ETHICS 5 (2014), http://www.i-r-i-e.net/issue21.htm.

[33] Tristan Denley, Advising by Algorithm, N.Y. TIMES (July 18, 2012), http://www.nytimes.com/interactive/2012/07/18/education/edlife/student-advising-by-algorithm.html?_r=0; Marc Parry, Colleges Awakening to the Opportunities of Data Mining, N.Y. TIMES (July 18, 2012), http://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html; Jeffrey R. Young, What the Slowdown in Ed-Tech Investment Means for Colleges, THE CHRON. OF HIGHER EDUC. (May 31, 2016), http://www.chronicle.com/article/What-the-Slowdown-in-Ed-Tech/236648/; Press Release, Ariz. State Univ., New Initiatives Advance ASU's Efforts to Enhance Student Success, https://asunow.asu.edu/content/new-initiatives-advance-asus-efforts-enhance-student-success.

[34] Long & Siemens, supra note 22.

student engagement. [35] Other HEIs capture RFID signals from RFID-embedded student IDs to track student movements to record course attendance, but also to analyze campus space and resource usage to correlate with academic performance.[36] Reading sensor data from RFID-embedded student IDs and analyzing wireless network traffic can also reveal social network connections among students. This opens up new doors for LA. Some proponents maintain that social network analysis can reveal student relationships, show the degree of cohesion within communities, and make clear what types of student relationships aid academic achievement.[37]

One novel means of information collection is being done by Oral Roberts University (ORU). ORU is now requiring its incoming students to purchase and wear Fitbit devices. Some Fitbit models can measure the number of steps an individual walks, her heart rate, quality of sleep, steps climbed, calories burned, and GPS coordinates throughout the day. ORU President William Wilson argues that tracking his students' physical fitness via the data streamed from the device is a part of the institution's mission to enhance each student's "mind, body and spirit."[38] The institution requires students to perform 10,000 steps a day and accomplish 150 active minutes a week, which the Fitbit records and feeds into the learning management system for instructors to review and grade.[39]

---

[35] Stan Alcorn, Facial Recognition in theIn The Classroom Tells Teachers whenWhen Students areAre Spacing, CO.EXIST (Oct. 18, 2013), https://www.fastcoexist.com/3018861/facial-recognition-in-the-classroom-tells-teachers-when-students-are-spacing; Ben Schiller, It'll be a lotBe A Lot Harder to CutTo Cut Class with thisWith This Classroom Facial-Recognition App, CO.EXIST (Feb. 17, 2015), https://www.fastcoexist.com/3042445/itll-be-a-lot-harder-to-cut-class-with-this-classroom-facial-recognition-app.

[36] David Brazy, *Arizona College college to Position Sensorsposition sensors to Check Class Attendance,check class attendance* THE BADGER HERALD (May 4, 2010), https://badgerherald.com/news/2010/05/04/ariz-college-to-posi/; Mary Catherine O'Connor, *Northern Arizona University to useUse Existing RFID Student Cards for Attendance Tracking*, RFID JOURNAL (May 24, 2010), http://www.rfidjournal.com/articles/view?7628 (last visited Nov 14, 2016).

[37] Daniel Amo Filvà et al,, *Social Network Analysis Approaches for Social Learning Support*, Proc.PROCEEDINGS OF THE SECOND INT'L CONF. ON TECHNOLOGICAL ECOSYSTEMS FOR ENHANCING MULTICULTURALITY (Oct. 1, 2014), http://doi.acm.org/10.1145/2669711.2669910; Parry, *supra* note 33; Simon Buckingham Shum & Rebecca Ferguson, *Social Learning Analytics*, 15 J. EDUC. TECHNOL. SOC. 3 (2012).

[38] Press Release,Univ.,*Integrates Wearable TechnologyPhysical Fitness CurriculumIncoming Students* (Jan. 4, 2016) Press Release, Oral Roberts Univ., Oral Roberts Univ. Integrates Wearable Tech. with Physical Fitness Curriculum for Incoming Students (Jan. 4, 2016) http://www.oru.edu/news/oru_news/20160104_fitbit_tracking.php.

[39] Jessica Chasmar, *Oklahoma University Requires Freshmenrequires freshmen to Wearwear Fitbit, Tracktrack 10K Stepssteps per day*, THE WASHINGTON TIMES, Jan. 11, 2016, http://www.washingtontimes.com/news/2016/jan/11/oklahoma-university-requires-freshmen-to-wear-fitb/; /Jeff Stone, *Not allAll Oral Roberts Students Need toTo Wear Fitbits, And They're Not Tracked Through Campus*, INT'L. BUS. TIMES, Feb. 3, 2016, http://www.ibtimes.com/not-all-oral-

Values at stake in higher education analytics

In the previous section we outline a number of ways in which higher education analytics is being used by HEIs. Many of the uses are at least prima facie justifiable. Colleges and universities should certainly expend resources helping ensure that they are providing ample educational opportunity for students. That includes knowing more about what is successful in classes, in advising, and in professional preparation. But what is prima facie justifiable is different from what is justifiable, all things considered. In this section, we examine a number of factors that may place moral limits on the scope of higher education analytics.

What is it for?

In its report on the state of higher education analytics, discussed above, the New America Foundation contrasts uses of analytics at Mount St. Mary's University (MSMU) and Georgia State University (GSU). The implication of the discussion is that while GSU's uses were admirable, MSMU's uses were not. Because MSMU sought to push people to leave in order to boost retention rates, at least as measured by reporting agencies. In Contrast, GSU sought to intervene in order to help students in danger of dropping out. It is worth considering, though, whether and why the one approach is better than the other. There are at least a couple of reasons.

First, GSU's approach advances the interests of the university's current students; in other words, the students whose information is monitored are the ones who benefit. Although there was a long period in which the university studied academic performance of all students before it started using information to intervene in its advising practices, the primary beneficiaries of GSU's approach are students. The ones who struggle, or the ones who appear to be at risk of doing poorly, get university help to advance goals that the students have set for themselves (viz., a college education). Contrast this with the MSMU's case. There, the school's strategy was to nudge students out. Now, one might argue that *if* a student is indeed unlikely to succeed in college, it is no benefit to allow them to continue paying tuition and forego other opportunities. That is true, but the key difference is that the MSMU case would have treated students who are at risk *absent further help* from a university as liabilities, whereas GSU treats those students as potentially successful.

Second, and related, is that data analytics *alone* does nothing whatsoever to advance anything of value. Instead, for analytics to be valuable, actions must be based on decisions about what is important, and resources must be allocated to advance those important things. In the MSMU case, the goal appeared to be better retention rates and, hence,

---

roberts-students-need-wear-fitbits-theyre-not-tracked-through-campus-2291808.

better national rankings. While that goal would appear artificial, after all, it is odd that a school's standing should be so influenced by something that can be easily manipulated and which is at best a proxy for something else, it could ultimately redound to the benefit of future students. But MSMU did not dedicate greater resources to achieving a worthwhile goal (e.g., investing in students in order to help them). Rather, it sought to game measures *to look like* it was achieving a worthwhile goal. In contrast, GSU's system was coupled with substantially increased resources for academic advising. It increased the number of academic advisors by hiring 42 new advisors, and its analytics system was geared to make use of those increased resources.[40]

Now, the main point here is that the goals at MSMU and GSU were superficially similar; each sought to advance institutional goals. But the key difference is that GSU sought to do so by advancing individual students' interests as well. This difference is worth emphasizing; institutional interests and student interests are not identical. They at times overlap, but they may come apart. It is crucial to be clear in any use of educational analytics just what interests are being pursued.

Other uses for information

In contrasting the GSU and MSMU's cases, we considered ostensible aims of analytics programs. But one of the baseline premises of data analytics is that by collecting lots of information from disparate sources, new and interesting correlations may be discovered. This raises the possibility that analytics systems developed with learning outcomes or even institutional goals in mind may come to serve very different purposes.

Consider first one of the key legal regimes addressing student privacy in higher education, the Federal Education Rights Protection Act (FERPA). Under FERPA, an "educational record" includes "records, files, documents, and other materials" directly related to an identifiable student who is or has been enrolled in a HEI.[41] The educational records must be maintained by the institution or other school officials, such as parties with whom the institution has contracted for services. There are some notable exceptions to what counts as educational records, including sole possession materials (e.g., records held by an individual instructor and shared with no one else), law enforcement records, employment files, and healthcare records in the possession of a health provider.[42]

Institutional actors and other school officials who have a "legitimate

---

[40] New America Foundation, *supra* note 2, at 2.
[41] 20 U.S.C. § 1232g (a)(4)(A) (2013).
[42] 20 U.S.C. § 1232g (a)(4)(B) (2013).

educational interest" in the educational record may access it without a student's express permission. [43] [44] But the definition of a legitimate educational interest is unclear. The Department of Education (ED) suggests that "a school official generally has a legitimate educational interest if the official needs to review an education record in order to fulfill his or her professional responsibility."[45]In addition to the leeway institutional actors have to disclose educational records within institutional boundaries, FERPA also enables disclosure to third-party actors–again, without student permission–under certain circumstances.[46]

Important for our purposes here, under FERPA, students have the right to inspect, review, and request amendments to their educational records.[47] That right to inspect is key, and something the audience for this symposium should be familiar with. HEIs may not release student records—for example transcripts—without student permission. That permission is what allows law schools, for example, to review applicants' transcripts. In an earlier paper, we have argued that the digital dossiers compiled on students in the context of LA may eventually become something that employers or others will begin requesting of applicants.[48] Proponents of educational analytics maintain that universities will be able to make more and better inferences about student performance and chance of success with more data and more sophisticated analysis. If LA lives up to that promise, it is easy to imagine others (e.g., future employers, credit reporting agencies) wanting access as well. If students can voluntarily allow access, those other parties may well seek permission.

Another issue to consider is the degree to which increased collection of student information could be turned to law enforcement or for security purposes. Consider section 215, the "business records" provision, of the USA Patriot Act. It allows the FBI to obtain a court order requiring other entities to produce "any tangible thing"—including any records—in order to protect against international terrorism. Specifically, it provides that

> the Director of the Federal Bureau of Investigation or a designee of the Director…may make an application for an order requiring the production of any tangible things (including books, records, papers, documents, and other items) for an investigation to obtain foreign intelligence information…to protect against international terrorism

---

[43] 20 U.S.C. § 1232g (b)(1)(A) (2013).

[44] *Id.*

[45] The Family Educational Rights and Privacy Act, Guidance for Eligible *Student*, DEP'T. EDUC. (Students, department of education, February 2011, last modified June 26, 2016),2015 – Rule 19. https://www2.ed.gov/policy/gen/guid/fpco/ferpa/for-eligible-students.pdf.

[46] 20 U.S.C. § 1232g (b)(2013).

[47] 20 U.S.C. § 1232g (a)(1)(2013).

[48] Alan Rubel & Kyle M. L. Jones, *Student Privacy in Learning Analytics: An Information Ethics Perspective*, 32 INF. SOC. 143–159, 150 (2016).

or clandestine intelligence activities….[49]

There are several limitations to the FBI's ability to collect information under section 215. Most importantly, there must be some grounds for believing that the items and records sought are "relevant to an authorized investigation."[50] Further, the FBI may not conduct an investigation based "solely" on activities protected by the First Amendment of the U.S. Constitution.[51] Moreover, the FBI must follow "minimization procedures" that limit the extent to which tangible things can be retained, disseminated, and used.[52]

Nonetheless, and despite these limitations, section 215 was the basis for the bulk telephone metadata collection program famously revealed by Edward Snowden.[53] And a well-integrated system of educational records about students—many of whom are citizens of other countries—might be an attractive target for security agencies.

Perhaps this is far-fetched and information about the comings-and-goings, friend and acquaintance groups, and other habits would be of little value or interest to security agencies. We (the authors) have no way of knowing. But the broader point is that it is merely an assumption that the ostensible or primary uses of data analytics will be the only uses. If the data analytics lives up to its promise, it is plausible that it will be put to other uses.

Relation to higher education mission

A question related to the purpose of higher education analytics concerns the underlying rationale for higher education in the first place. In our discussion above, we consider different uses of analytics, among them are advancing learning outcomes, operational efficiency, and student retention. Each of these seems like a good thing, considered in isolation. However, it is important to consider what the aim of higher education is. In Rubel and Jones, in a previous article we consider one important conception of higher education, advanced by Derek Bok.[54] Bok argues that universities

---

[49] 50 U.S.C. § 1861(a)(1) (2015).

[50] 50 U.S.C. § 1861(b)(2)(A) (2015).

[51] 50 U.S.C. § 1861(a)(1)-(2) (2015).

[52] 50 U.S.C. § 1861(b)(2)(B) (2015).

[53] Ewen MacAskill, *The NSA's Bulk Metadata Collection Authority Just Expired. What now?*, THE GUARDIAN (Nov. 28, 2015) https://www.theguardian.com/us-news/2015/nov/28/nsa-bulk-metadata-collection-expires-usa-freedom-act.

[54] Rubel and Jones, *supra* note 45, at 155; DEREK CURTIS BOK, OUR UNDERACHIEVING COLLEGES: A CANDID LOOK AT HOW MUCH STUDENTS LEARN AND WHY THEY SHOULD BE LEARNING MORE (2006).

and colleges should aim at improving students' capacities for communication, critical thinking, citizenship, living in a diverse-global society, and employment.55 We argue that each of these goods is at least partially constituted by persons' *autonomy*. That is, the ability to communicate, think critically, and be a citizen in an increasingly diverse and global society demands that one be able to think for oneself about what matters and to act according to what one's values as one sees fit. However, if universities seek to track students ever more carefully, and to intervene in students' decisions about educational goals, they act contrary to students' abilities to act for their own reasons. Recall that one of the flags that GSU uses to intervene with students is when they take courses not needed for their majors.56 We have no reason to doubt that taking courses not needed for one's major is correlated with a risk of performing poorly. However, it is also a concern if universities seek to nudge students away from expanding their intellectual reach. And indeed, one of the primary moral justifications for privacy is the ability to make decisions for oneself, without undue influence from others.57

II.   A RESEARCH AGENDA

So far, we have outlined some important features of data analytics in higher education, and we have suggested a few key concerns (what purposes do particular analytics programs serve, how else will such programs be used, and what ultimate aims HEIs should reflect). But there is a great deal more to be done to understand the impact of data analytics in higher education. In this section, we outline several key areas that warrant further, long-term observation and analysis.

The first big set of questions concerns *other actors*. The focus of the learning analytics literature and practice has until now been on students and advisors: what student behaviors and actions correlate with better outcomes (e.g., grades, retention). But there are other actors involved as well—instructors, for example. Course management systems (e.g., Blackboard, Moodle, Desire2Learn) record the actions of all people using the system. It is not implausible to think that administrators will begin to track more closely how instructors operate within courses: does greater time spent reviewing student work correlate with better student performance? Does greater interaction with students on discussion boards lead to increased

---

55 BOK, *supra* note 50, at 66.

56 New America Foundation, *supra* note 2, at 3.

57 STANLEY I. BENN, PRIVACY, FREEDOM, AND RESPECT FOR PERSONS, IN PRIVACY: NOMOS XIII 1–26 PRIVACY: NOMOS XIII 1–26 (J. Roland Pennock & John W Chapman at al eds., 1971) (RULE 15.5.1); Jeffrey H. Reiman, *Privacy, Intimacy, and Personhood*, 6 PHILOS. PUBLIC AFF. 26–44 (1976); Alan Rubel, *Privacy and Positive Intellectual Freedom*, 45 J. SOC. PHILOS. 390–407 (2014).

enthusiasm and retention rates? And will job performance be judged on these measures (which may be more easily measurable) rather than others that may be more difficult to capture (creativity, challenging material, pushing students to develop their independent thinking rather than mastering selected material)?

Another aspect of data analytics has less to do with data gathering and analysis, and more to do with who will control the environment in which instruction takes place. For analytics to be most effective it demands some degree of standardization and increasing volumes of data. As noted above, a great deal of the information used in learning analytics is gathered in course management systems. However, instructors may wish not to use those systems (perhaps preferring materials on paper and grades on spreadsheets). It is plausible that HEIs will push instructors to instead use CMS in order to better capture student information. Similarly, institutions may prefer that instructors use *one particular* CMS in order to better collect and share information.

But why should this matter? After all, instructors are employees of universities and universities can control at least some aspects of their employment. One possibility is that instructors have a claim to academic freedom, which includes the ability to pursue research and teach in the manner that they deem best. That does not mean that instructors can do any old thing in the classroom, but it may mean that they should control the basic structures of courses they teach. And, hence, should be able to choose whether, and what type, of course management system to use, and should be able to manage their classes in ways that they consider optimal for instruction even if they are not optimal from the standpoint of data analytics. Another possibility is that pushing instructors to use particular CMS will prevent them from doing the kind of experimentation that leads to educational innovation.

Now, there is an important body of literature about academic freedom in higher education, and that literature represents a wide range of views about the nature and scope of academic freedom. And understanding that literature will be an important part of addressing conflicts between institutions and instructors on the subject of data collection and analysis about instructors.

Another issue regarding other actors concerns *vendors*. The attraction of Big Data and the potential for profits has led to "explosive activity" around LA, and within HEIs and the educational technology marketplace.58 EDUCAUSE's *Horizon Report* has described increasing interest in and

---

58 Ronald Yanosky & Pam Arroway, *The Analytics Landscape in Higher Education*, 24, EDUCAUSE Center for Analysis and Research 2015, https://library.educause.edu/~/media/files/library/2015/5/ers1504cl.pdf.

adoption of LA each year since 2010; each annual issue highlights the increasing maturity of the technology and institutional ability to use LA to further educational, management, and other. In the first quarter of 2014 alone, educational technology startups raised over $500 million from venture capitalists, which is an increase of over $200 million in the previous year.[59] Recent reports indicate that investments in educational technology have slowed in 2016, but LA remains a primary driver of investments; moreover, LA is predicted to grow up 25% by 2020.[60]

There have been notable acquisitions and venture capital fundraising among some of the top educational technology companies offering LA applications and systems to the education institutions. Desire2Learn is an illustrative case. In 2012, the company raised $80 million dollars in funding; at the time, according to Desire2Learn, the influx of cash was the largest ever investment in a Canada-based software company.[61] One of the motivations behind the investment was a need to hire developers to build up their LA toolset.[62] Additionally, the company went on a buying spree, acquiring Degree Compass and Knowillage, among other companies, to support their LA initiatives.[63] In the same vein, Civitas Learning raised $60 million to further establish LA as the "backbone" of its data-driven systems;[64] Google's funding arm invested $40 million dollars in Wisconsin-based Renaissance Learning;[65] and Blackboard acquired X-Ray Analytics, a predictive LA company, to enhance their understanding of learner behaviors at "course, cross-course and institutional levels."[66]

The relationship between educational technology start-ups, institutions,

---

[59] Jonathan Shieber, *Education Technology Startups Raised Over Half A Billion Dollars In Q1*, TECHCRUNCH (March 26, 2014), http://social.techcrunch.com/2014/03/26/education-technology-startups-raised-nearly-half-a-billion-dollars-in-q1/./

[60] Bothwell *supra* note 12 Ellie; Young, *supra* note 33.

[61] Press Release, Desire2Learn, Raises $80 Million in Financing Round from NEA and OMERS Ventures (Sept. 4, 2012) https://www.d2l.com/newsroom/releases/desire2learn-raises-80-million-in-financing-round-from-nea-and-omers-ventures/; Tony Wan, *Desire2Learn's $85 Million Deal,* EDSURGE NEWS, (Aug. 13, 2014), https://www.edsurge.com/news/2014-08-13-desire2learn-s-85-million-deal.

[62] Ki Mae Heussner, *As ed Tech Heatstech heats up, Desire2Learn Raisesraises $80M in its Firstfirst VC Round,round* GIGAOM (Sept. 4, 2012, 9:48 AM), https://gigaom.com/2012/09/04/as-ed-tech-heats-up-desire2learn-raises-80m-in-its-first-vc-round/. /

[63] Tony Wan, *The Sum of Desire2Learn's Acquisitions: Brightspace*, EDSURGE NEWS (July 14, 2014), https://www.edsurge.com/news/2014-07-14-the-sum-of-desire2learn-s-acquisitions-brightspace.

[64] Shieber, *supra* note 55.

[65] Frederic Lardinois, *Google Capital Invests $40M inIn Learning Analytics Firm Renaissance Learning At $1B Valuation*, TECHCRUNCH (February 19, 2014), http://social.techcrunch.com/2014/02/19/google-capital-invests-40m-in-learning-analytics-firm-renaissance-learning-at-1b-valuation/./

[66] *Blackboard's Buying Spree Continues: X Ray Analytics Becomes 10th Acquisition Since 2014*, EDSURGE NEWS (July 6, 2015),EDSURGE https://www.edsurge.com/news/2015-07-06-blackboard-s-buying-spree-continues-x-ray-analytics-becomes-10th-acquisition-since-2014 (last visited Nov 14, 2016).

and the public has been a tenuous one, regardless of the money investors have put into developing LA. Consider the demise of inBloom, a company that sought to aggregate and store student information on behalf of states and K-12 school districts. In addition to acting as a data steward, inBloom offered third parties access to the data warehouse to enable application and dashboard development. Parents, guardians, and legislators cried foul. After receiving $100 million in seed money from the Bill and Melinda Gates Foundation, the company shut down after prolonged fights with concerned stakeholders about privacy, especially related to the sensitivity of the data inBloom collected and made accessible. [67] The databases included, among other details, each student's family situation (e.g., that the student's father had a "significant other"), enrollment changes (e.g., whether the student left a school due to a violent incident), and whether or not a student had a learning disability.

In the wake of inBloom's failure, advocates and educational technology industry leaders alike came together to commit to protect student privacy.[68] Specifically, the alliance between the Future of Privacy Forum (FPF) and the Software & Information Industry Association (SIIA) resulted in the Student Privacy Pledge. Jules Polonetsky, executive director and co-chair of the FPF argued that "the Pledge will enhance the trust between families, schools and third party service providers" and hold providers to a higher standard, setting them apart from their industry peers as ethical actors.[69] The Pledge requires those who sign it to adhere to a set of key principles, which together require them to:

> Not sell student information
> Not behaviorally target advertising
> Use data for authorized education purposes only
> Not change privacy policies without notice and choice
> Enforce strict limits on data retention
> Support parental access to, and correction of errors in, their children's information

---

[67] Natasha Singer, *Senator Raises Questions About Protecting Student Data*, N.Y. TIMES BIT BLOG (Oct. 22, 2013, 12:01 PM), http://bits.blogs.nytimes.com/2013/10/22/senator-raises-questions-about-protecting-student-data/;/ Natasha Singer, *InBloom Student Data Repository to Close*, N.Y. TIMES BIT BLOG: (April 21, 2014), http://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close.

[68] Press Release, Student Privacy Pledge,).Original Press Releasepress release for the Pledgepledge to Safeguard Student Privacy,safeguard student privacy https://studentprivacypledge.org/official-press-release/. /

[69] *Ibid*.

As of November 2016, over 300 companies had signed the Pledge.[70] The issue here is that higher education institutions and their students, staff, and constituents, are far from the only parties with an interest in data analytics. Moreover, the fact that there are political pressures to pursue analytics means that there is potential for substantial profit. This raises several important questions with respect to vendors. First, will vendors be able to capture HEIs and ensure long-term use of their systems and, hence, a substantial revenue stream? If so, will that be because of good products that advance institutional ends, or will it be due to political pressures to employ data analytics? Further, will HEIs retain rights to their own data, or will vendors own rights to the data in order to sell it back to universities (much in the way that academic publishers sell journals to university libraries, even though university personnel write, review, and edit the contents of the journals).

The basic structure of LA rests upon substantial efforts by HEIs, work by instructional and information technology staff, and buy-in (at least tacitly) from students. In other words, HEIs expend resources and incur opportunity costs to tailor their information systems, including their course management systems, in such a way to gather data in a comprehensive and useful way. Moreover, the information systems that HEIs use are developed by third parties, and there is substantial money to be made in selling information systems and analytic tools to HEIs and consortia. One important concern, though, is that LA not fall into the cycle of scholarly publishing.

The traditional model of scholarly publishing is roughly as follows. HEIs and granting agencies (often public, such as National Science Foundation, National Institutes of Health, National Endowment for the Humanities, and Institute for Museum and Library Studies) fund research, which is carried out by scholars (faculty, research staff, post-docs, students). The researchers analyze prior, related work, and compile and analyze their novel contributions into manuscripts. They share those manuscripts with publishers, who send the manuscripts to other scholars and researchers for peer-review (also subsidized by HEIs and granting agencies). When a publisher agrees to publish the manuscript, the authors/researchers assign the copyright to the publisher. The publisher covers costs of copy editing, typesetting, marking up in HTML, printing, distribution, and marketing, and in turn licenses the manuscript (collected along with others) to libraries, which are funded by HEIs and other granting agencies. In turn, libraries make the manuscripts available to researchers,

---

[70] *K-12 School Service Provider Pledge to Safeguard Student Privacy,* STUDENT PRIVACY PLEDGE (Oct. 7 2014), https://studentprivacypledge.org/wp-content/uploads/2014/09/Student-Privacy-Pledge-V1.pdf. Student Privacy Pledge. (2015).

scholars, and students who use the material as a basis for further research.[71]

There is potential for higher education data and analytics to fall into the same cycle. Universities and consortia already enter into agreements with third parties to gather, analyze, and repackage education data (e.g., D2L, Blackboard). This mimics the first part of the scholarly publishing cycle, in which HEIs subsidize research that is made available to third parties. Those third parties can develop analytics tools that they can then license back to universities. If universities were to assign exclusive rights to the data to those third parties (as has happened in scholarly communications), those third parties can extract substantial fees from universities in order to use the analytic tools. Moreover, because of the political pressures to make university policy more data-driven (for good reasons, such as student retention and learning outcomes, or for more suspect ones, such as allocating department support based on weak metrics) university demand might be inelastic. With a few firms dominating the landscape, it is not hard to imagine higher education analytics closely resembling the scholarly publishing model.

The above research agenda considers what is likely to happen in data analytics: how will different actors be affected, and how will vendors affect the function and sustainability of higher education institutions. But there's a separate research question that is at root about value. Information collection, analysis, and use may conflict with at least some conceptions of the value of higher education. But there is ample opportunity to consider different possibilities of the value of higher education.

A recent volume edited by Harry Brighouse and Michael McPherson collects recent scholarship about the proper ends of higher education.[72] The breadth of views articulated shows that precisely what justifies the existence of higher education institutions is far from settled. So, for example, a number of commentators consider HEIs in terms of *economic* impact, either as contributors to aggregate economic growth or means of more justly distributing economic goods.[73] Amy Gutmann's account in the same volume outlines several compelling goals. First is that higher education should provide opportunity on the basis of talent and work (rather than, e.g., birth and wealth). Second is that higher education should aim at "greater integration of knowledge…within the liberal arts and sciences [and] between liberal arts and professional education."[74] This she refers to

---

[71] PEGGY JOHNSON, FUNDAMENTALS OF COLLECTION DEVELOPMENT AND MANAGEMENT (3d ed. 2014); JOHN J. REGAZZI, SCHOLARLY COMMUNICATIONS: A HISTORY FROM CONTENT AS KING TO CONTENT AS KINGMAKER (2015).

[72] HARRY BRIGHOUSE & MICHAEL MCPHERSON, THE AIMS OF HIGHER EDUCATION: PROBLEMS OF MORALITY AND JUSTICE (eds., 2015).

[73] *Id.* at 26-51; 135-55.

[74] *Id.* at 7-8.

as creative understanding. Third, undergraduate education should foster community engagement and contribution based on their creative understandings.75

What is important in each of these is that they will provide the basis (if any) for substantial investment in data analytics and the rationale (if any) for the justifiable *uses* of learning analytics. If, for example, higher education institutions are justified on the basis of economic impact, analytics should reflect that. If instead higher education is justified on the grounds community engagement and creative understanding, then perhaps very different measures and goals should ground our institutions' practices.

Hence, an important part of a research agenda for data analytics in higher education will be to better understand what ground higher education in the first place.

III. CONCLUSION

Data analytics is everywhere, but the participants in this symposium are intimately involved with higher education analytics; we see it every day: we use it, we are affected by it, and (most importantly) we have a better chance of holding some sway over it. Our task here has been to describe some of the basic features of higher education analytics on the ground today, outline some values at stake in the enterprise (it's purposes, other ways analytics might be used, how it relates to higher education's value), and chart a research agenda for better understanding how data analytics will unfold in the higher education landscape (how will it affect other actors, how will vendor influence affect HEIs, what is the best understanding the value of higher education and how would that be advanced by analytics). There's much there to consider, and more work to be done. But it is certainly worth asking of our own institutions just what goals we wish to advance and how analytics advances them.

---

75 *Id.*