Student Privacy in Learning Analytics: An Information Ethics Perspective

Alan Rubel
Assistant Professor
School of Library and Information Studies
Program in Legal Studies
University of Wisconsin-Madison
arubel@wisc.edu

Kyle Jones
School of Library and Information Studies
University of Wisconsin-Madison
Kmjones8@wisc.edu

1. Introduction

Commercial enterprises have been at the forefront of data analytics. They have collected enormous amounts of data about consumers and their behavior, developed tools to analyze that data, and used the results to refine their business practices. In recent years, educational institutions have started using similar tools in higher education. By gathering information about students as they navigate campus information systems, learning analytics "uses analytic techniques to help target instructional, curricular, and support resources" to examine student learning behaviors and change students' learning environments (van Barneveld, Arnold, & Campbell, 2012, p. 8). As a result, the information educators and educational institutions have at their disposal is no longer demarcated by course content and assessments, and old boundaries between information used for assessment and information about how students live and work are blurring.

There is no question that learning analytics is potentially useful in higher education. Nonetheless, collecting and using information about students raises a number of moral questions, including issues pertaining to student privacy. And while the literature on learning analytics recognizes potential privacy conflicts, there is little systematic discussion of the ways in which privacy and learning analytics conflict. Our task here is to address this gap and provide a framework for understanding and evaluating privacy issues in the context of learning analytics.

We begin, in section 2, by describing the basic features of learning analytics, including current applications, data sources, uses, and initiatives. In order to better understand the current state of learning analytics we provide some background information regarding actors, goals, motivations, and incentives, and describe some important currents in the evolution of higher education, educational technologies, and student populations. We then turn to some important moral and policy questions, focusing on issues surrounding student privacy. After setting out some basics about the nature and value of privacy generally, we argue that there are five crucial questions about student privacy that we

Draft: Please cite to final version, forthcoming in *The Information Society*.

must address in order to ensure that whatever the laudable goals and gains of learning analytics, they are commensurate with respecting students' privacy and associated rights, including (but not limited to) autonomy interests.

We group these questions into four *narrow* questions, and one *wide* question. The first narrow question is who all may access student information, and whether one entity's access makes others' access more likely. The collection and analysis of student information by an institution (for example) should not imply that students' privacy should diminish with respect to other entities, though we will argue that this seems to happen nonetheless. Second is whether some information, though useful, is too instrusive to warrant collection. We argue that we need clear criteria for what information may justifiably be collected in the name of learning analytics. Third, learning analytics is often justified on the grounds that it will lead to better consequences, namely better learning outcomes. However, determining whether such consequences do indeed justify collecting and analyzing large amounts of student information requires a careful accounting of those consequences, alternative paths that may lead to different sets of consequences, and the distributions of those consequences. Fourth, regardless of how robust the benefits of learning analytics turn out to be, there is a question as to whether those benefits are sufficient to outweigh students' important autonomy interests in how information about them is collected. The wide question is whether the goods that justify higher education are advanced by learning analytics, or whether collection of information actually runs counter to those goods.

2. Background: Learning Analytics

What is learning analytics?

Learning analytics is the collection, analysis, and use of large amounts of student data and information to better understand learner behaviors and contexts (both digital and analog) to improve learning outcomes and to increase institutional efficiency and effectiveness (Siemens, 2013; van Barneveld et al., 2012). In 2010, the annual *Horizon Report* identified the use of advanced computational methods and data visualization techniques (core technological components of learning analytics) as an important strategy for understanding student learning (Johnson, Levine, Smith, & Stone, 2010). Since then, interest in learning analytics has grown and technological capabilities have increased. Learning analytics is not limited to particular technologies and methods. Rather, it is a technology-enhanced practice that encompasses many different tools and techniques to analyze various types and amounts of student data. It relies on the "digital breadcrumbs" students leave as they interact within information systems on and off-campus (EDUCAUSE Learning Initiative, 2011). Such breadcrumbs include records of student logins and logouts, maps of student clickstreams, timestamps of activities and resource access, and any text inputs (e.g., discussion forum posts) students provide within information systems.

In contrast to big data practices in commercial and research domains, where de-identified data may be useful, learning analytics links digital breadcrumbs to individual students. Individual linkage is key for analytics to be most useful to institutions and students. In other words, for learning analytics to be useful for individual students, data must retain unique identifiers.

The potential scope of information useful for learning analytics is unlimited. Proponents of learning analytics attempt to gather "any data [they] can get their hands on" (Cite redacted, in progress) to uncover hidden relationships in a mass of data using statistical relationships, and to use predictive tools to "foresee events before they happen" (Mayer-Schönberger & Cukier, 2013, p. 58). Other techniques useful for learning analytics include use of social network analysis to map student engagement in a course and adaptive content to support individualized learning (Long & Siemens, 2011). Another useful tool is data visualization; once the statistical calculations have been made, the analytical findings are often displayed in graphical user "dashboards," so students, instructors, and administrators alike can visualize the data and employ it as actionable information.

The development of learning analytics

Learning analytics' rise in prominence is due in large part to the data infrastructure created by online learning initiatives, outside pressures for higher education reform, and the use of data in support of each. The evolution of this data infrastructure can be seen in three waves (Brown, 2011, p. 2). The first is the wide adoption of learning management systems (LMS) to host and administer online courses, the creation of repositories of learning materials, and an increasing population of online learners using LMSs. To illustrate, over 6.7 million students enrolled in at least one online course in 2012 and the overall proportion of all students taking an online course was at a high of 32% in 2013 (Allen & Seaman, 2013, p. 4). The second wave is related to social learning applications, or layers, being added to the LMS to allow for networked learning opportunities among students. These two waves have led to a "data explosion" (Long & Siemens, 2011, p. 32), which has created the necessary data conditions for analytic technologies. The most recent, third wave is the use of this data, including mining data to develop strategies for instruction, advising, infrastructure, and resource allocation (Brown, 2011, p. 2).

In addition to the LMS, higher education institutions commonly employ other information systems. For example, student information systems (SISs) store each student's socio-economic status, demographic profile, academic history, and financial aid package. Libraries use integrated library systems (ILSs) and related technologies to track material borrowing and capture what digital resources students access. As more courses use electronic textbooks, campuses can capture student reading habits (see Alter, 2012; Dennis, Duffy, & Morrone, 2010). And campuses often use single sign-on systems (SSOSs) for campus applications and networks, which have the capacity to store unique pieces of data, which are either input directly by the student or captured as students interact with a system. Together these data-based systems create a web of information systems that captures a significant amount of data *about* students and data generated *by* students.

Another force driving learning analytics is outside social pressure on educational institutions to make substantive reforms and "prove" their success with data (Wagner & Ice, 2012). Data-driven higher education, the argument goes, could increase transparency about an institution's inputs, outcomes, and processes in order to spark improvement. For example, more visible data about student retention levels and expectations may help stakeholders (potential students, enrolled students, parents, accreditation bodies, and legislators) to evaluate schools and programs.

Influential non-profits and the federal government have also championed the big data environment, hoping to change higher education with analytic techniques and data services. The Hewlett and Gates foundations granted nearly $24 million to higher education projects as part of their Next Generation Learning Challenges program; $3 million of that money went to learning analytics projects (Grantees, n.d.; Wave I, n.d.). The Obama administration's MyData initiative seeks to provide students with downloadable copies of their academic record–including online learning activity–to empower students and build up a market of third-party analytic services to mine the downloaded data for more value (Chopra & Smith, 2012; MyData, n.d.; Office of Science and Technology Policy, 2012). The administration has worked closely with companies like Pearson and Microsoft to develop interoperability standards between the MyData portal and their respective educational technologies to include larger sets of data for students to access.

In an effort to capture comprehensive sets of historical academic data, legislators have also renewed conversations about creating a federal unit record system (FURS) (Nelson, 2013). In effect, FURS would enable the Department of Education to track students' academic performance throughout their entire academic careers and record their wages afterwards. Connecting datasets from the primary, secondary, and post-secondary levels may boost the statistical power of learning analytics systems and inform curriculum design at the lower levels. It could also be included in the MyData project.

The direction of learning analytics

Although improved outcomes is one overarching goal, Long and Siemens (2011, p. 36) have discerned a number of goals learning analytics could be used for:

- better institutional decision making and resource use;
- improved learning for at-risk students;
- increased institutional transparency;
- transformative change to teaching methods;
- better insight into networked knowledge;
- data-driven experimentation for administrative problems (e.g., enrollment and retainment);
- increased "organizational productivity and effectiveness";
- value-ranking of faculty activity;
- comparative learning metrics for students (e.g., how a student compares to her peers in a particular area).

Illustrating these themes, Michael Crow, president of Arizona State University (ASU), says "there are no more excuses" for not putting analytics to work on sticky problems like enrollment, increasing the diversity of the campus's student body, and monitoring and communicating unusual behavior of individuals to increase the security of the campus and safety of its community (Oblinger, 2012, p. 21). Another example is Coppin State University (CSU), which has fully embraced a "culture of performance measurement and improvement" that learning analytics may bring about in higher education (Norris, Baer, Leonard, Pugliese, and Lefrere, 2008, p. 48). CSU used a combination of learning analytics and other institutional analytics to provide information and proof of success when its education school sought reaccreditation from the National Council for Accreditation of Teacher Education.

There are a number of ways in which learning analytics may develop in the coming years. One is by combining data sources. Learning analytics technologies source their data primarily from campus LMSs and SISs. Together, this merged data provides a foundation for learning analytics to make predictions of learning outcomes based on a student's personal, financial, and academic profile. Demonstrating this merge of data and its potential, the American Public University System (APUS) used a comprehensive data warehouse to run semantic analyses and statistical tests to rank its entire student body by probability of success, intervening with students whose rank was at the tail end (Norris, 2011). The University of Phoenix, an online, for-profit institution, used merged datasets to "predict the likelihood" of student failure and "prioritize students" for academic interventions, primarily based on demographic and financial information (Barber & Sharkey, 2012, p. 259). Similarly, Rio Salado College's (RSC) RioPACE system identified students whose learning behaviors (e.g., log-in frequency, pace of work, and engagement) might put them "at-risk" of not successfully completing a course (Grush, 2011; Norris, 2011).

Some colleges and universities have aggregated LMS and SIS-type data across institutions. In order to build predictive models about student retention and student progression, APUS, Colorado Community College System, RSC, the University of Hawaii System, University of Illinois-Springfield, and the University of Phoenix developed the Predictive Analytics Reporting (PAR) framework (Ice et al., 2012). The aggregated dataset amounted to over 3 million course records and 640,000 student records, which were de-identified due to the experimental nature of the project.

Another purpose for using learning analytics is to intervene when students are performing poorly. For example, some researchers argue that learning analytics should employ nudging techniques (i.e., scripted and manual intervention strategies) to encourage positive behaviors (Carmean & Mizzi, 2010). The work done by Arnold (2010) and Purdue University's Course Signals system, for example, uses a red, yellow, and green light system to nudge students to access resources to improve in their course. Instructors can customize the system to push "e-mails and reminders, text messages" and more automated, system-generated interventions directly to students (Mattingly, Rice, & Berge, 2012, p. 243).

The University of Texas, ASU, and Harvard University have begun to develop their own adaptive analytics technology, which adjusts course content for individual learners based on their learning behaviors, past performance, and system-identified learning styles (Parry, 2011; Parry, 2012). At Harvard University, their "Learning Catalytics" software even analyzes differences in student responses to in-class questions to pair students up in real-time in a course for class discussion (Parry, 2011).

Institutions are also developing automated eAdvising systems to guide (and track) students through the process of choosing courses for their majors. Austin Peay State University uses recommendation algorithms to suggest courses based on a student's chosen major, requirements for graduation, and academic performance in relationship to that of her peers (Denley, 2012). ASU's eAdvising system intervenes by e-mailing students and their advisors, displaying messages on students' advising dashboards that they are “off-track” from their system-generated progress plan, and blocking students from registering for courses if they fail to take action (ASU, 2011; Parry, 2012).

Another potential development is that learning analytics may expand into non-education data sources. Pioneering work in learning analytics is moving towards capturing data that students-as-users create on the social web. “Every Tweet, every Facebook status update, every social interaction,” argue Long and Siemens (2011, p. 32), provides an opportunity for institutions and researchers to understand learning behaviors. With this in mind, Samford University is mining social and behavioral data patterns from users in their “Class of 2017” Facebook group, which every admitted student for 2013 was officially invited to by the institution, in order to discover “who’s likely to enroll” and “who’s on the fence” (Hoover, 2012, para. 2). While this has less to do with learning and more to do with admissions, it takes only a small conceptual leap to envision how students who actually enroll are continually tracked in Facebook and how that data could inform services and offerings related directly to their learning.

There are further opportunities to tap other rich information sources. Geolocation data presents an especially rich opportunity for institutions to expand their data points. Institutions commonly implant student IDs with RFID chips. The infrastructure already exists to passively reads the IDs when students present them to enter into campus buildings; the same infrastructure could be enhanced to actively read the IDs without students necessarily interacting with the ID reader. It is plausible to think that institutions would install active readers at the entrances to campus buildings, at the doors of their transportation services (e.g., on their buses), and at strategic points around campus critical to student learning (e.g., library learning labs). Similar geolocation data could be extracted from SSOSs by collecting the Internet Protocol (IP) address of students or by matching students to specific wireless hotspots on campus when they login.

Student tracking throughout campus infrastructures enables an institution to make hypotheses about factors that could enhance their learning. One could argue that RFID-enabled tracking provides insight into high-use and low-use campus resources by a student, which would provide data for learning analytics systems to nudge students to, say, use the library more often; or, as Matthew Pattinsky of the LMS company Blackboard argues, to map out a campus’s social network and pinpoint student engagement in the life of the university (Parry, 2012).

Some institutions may value data that reveals information regarding the welfare of its students, especially their health. Campus wellness centers already keep medical records about their students and student unions use student IDs to record food purchases, so it is certainly plausible that health data could be folded into the algorithm to inform advisors or instructors why, perhaps, their student may not be succeeding in a course.

Furthermore, Samford University and other institutions could expand their use of Facebook profile data to tie the information in it more directly to the institution. Some students may find it extremely helpful, for example, if their institution’s analytics system matched their self-ascribed hobbies, predilections, and views to services and resources: a student may express her belief in Judaism and the system could automatically provide them with contact information (e.g., phone, e-mail, and

Facebook page) to the campus's Center for Jewish Studies and Jewish student organizations. Or, to expand this idea further, the analytics system could run discourse analyses on a student's profile updates in order to alert advisors of concerning behavior, such as suicidal tendencies (see Mandge, 2013 for work in this area).

Concerns

There are a number of potential problems about big data in general that apply to learning analytics. First, scholars have raised concerns about data-driven institutions and the increasing power they have over the individuals about whom they collect data, citing that the information mined from the data is exponentially more valuable for the institution than the data subject who increasingly has little recourse or knowledge in regard to managing her personal data (Jerome, 2013; Tene & Polonetsky, 2013). Second, big data practices aim to "make the world more transparent," but the data they subsume and the resulting analytical products that influence the lives of individuals are black boxed. Richards and King (2013) call this the "transparency paradox," and it creates a Kafka-like system that affords individuals little information regarding what data is gathered about them and to what ends it may be used. Third, big data practices classify individuals based on one's socioeconomic status, race, ethnicity, or gender–all common data classifications. And when predictive analytics are used without using other explanatory data with or instead of these sensitive data classifications, the result is that statistical models and their resulting action may continue to perpetuate "old prejudices" (Tene & Polonetsky, 2013, p. 254). Finally, as individuals become more aware of the existence of the mass of data about them and the purported and actual ends to which it has been put, they may consciously change their behaviors based on who or what is recording data about them; this is, of course, the result of the "chilling effect," which is often brought on by concerns about surveillance (Solove, 2006; Stanley, 2012).

2. Privacy

What is privacy?

Privacy is a contested concept, and there is a substantial philosophical literature debating what privacy is, whether privacy is valuable, and if it's valuable, why. We won't attempt to adjudicate between different conceptions of privacy here, but it is important to set out several key aspects of privacy that will be important in the context of learning analytics. And although we will use a particular conception of privacy, our arguments will be compatible with a broad range of views in the literature.

But just what does it mean for one to have privacy? To begin, a conception of privacy should be understood as a three-part relation between some person or persons P, some domain of information O, and some other person or persons Q, such that P has privacy regarding O with respect to Q (Rubel and Biava 2014; Rubel 2011; see also Blaauw 2013). Merely stating that "P has privacy" or "P lacks privacy" is incomplete and often too vague to address moral or policy questions about privacy. For example, if we are considering a question about student privacy, the claim that some student P "has privacy" will not be very helpful. Rather, we will need to specify what it is that P has privacy *about* (her grades? her medical history?), and we will need to specify who it is that P has privacy with respect to (the public?

Her school?). Hence, P may have privacy regarding her grades in her major (O) with respect to a prospective employer (Q1) while at the same time not have privacy regarding grades in major (O) with respect to her academic advisor (Q2). Keeping straight just what POQ relation is at issue will be important in when we begin addressing privacy claims, below.

Next, there is a question of whether a claim about privacy (that is, whether some P has privacy regarding O with respect to some Q) merely describes some state of affairs or instead entails some value claim about Q's access to information about P in domain O. For example, Judith DeCew argues that privacy applies to information that "is not generally—that is, according to a reasonable person under normal circumstances…a legitimate concern of others" (DeCew 1997: 58), hence building value into the very concept. Other accounts are merely descriptive, such that one can have diminished privacy in some domain with respect to some other person. It is a further question whether that privacy loss is morally problematic, and further question still whether one has some moral claim to maintaining that privacy (see Rubel 2011; Moore 2010: 26-27).

Another question addressed in the literature is what the privacy relation actually involves; that is, what it means to say that P's privacy regarding O with respect to Q decreases. There are a variety of views, including (among others) information control, knowledge, and inference (see, for example, DeCew 1997, pp. 53–54; Fried 1970, pp. 140–141; Parent 1983; Allen 1988, 15; Rubel 2011). Here we use an *access* account, according to which P's privacy regarding information in domain O is diminished with respect to Q where Q's ability to access information about P regarding O increases.

Helen Nissenbaum's contextual integrity approach seeks to account for these nuances by understanding privacy in terms of whether information flows are appropriate to specific contexts, or "structured social settings" constituted by actor roles, accepted and expected activities, behavior-guiding norms, and objective or ends-oriented values (Nissenbaum 2010: 134-135). Contexts determine to whom, from whom, and about whom information should flow; what the information is about (its attributes); and whether or not there exist "terms and conditions" of the information flow (or, transmission principles). Put together, contexts, actors, attributes, and transmission principles form information norms, and where some aspect of an informational context changes (e.g., a new technology, rule, practice, or the like), defying entrenched expectation there is a prima facie violation of contextual integrity. Hence, contextual integrity is a "benchmark for privacy," for determining whether there is a potential privacy problem (Nissenbaum 2010: 150). It is a further question, though, whether a prima facie violation of contextual integrity is morally problematic, all things considered.

Why is privacy valuable?

Even once we have explained what privacy is, there remains a question of whether or not and why privacy is valuable. A number of commentators have referred to privacy's instrumental value—it is valuable for what it does. Privacy in some domains, with respect to some others, may promote a variety of social relationships. Intimate relationships, for example, may do better where others are unable to access information important to those relationships. And as James Rachels (1975) has argued, it may be useful in having a variety of relationships important in life to limiting information within that relationship. Having arms-length business relationships is aided where the parties do not know too much about other facets of the others' lives, and (directly relevant to this paper) student-teacher relationships demand a degree of distance, which is in turn aided by limiting the flows of some information. Knowledge of how a student spends his free time could undermine a relationship with a teacher if the teacher were to not approve of the student's activities.

Others have described privacy as important as a function of autonomy. Autonomy is a complex concept, and there is substantial scholarly debate about what it means and the extent to which it is morally important. For our purposes here, though, a rough account will suffice. Autonomy includes having the ability to self-govern, that is, to be able to make decisions for oneself, based on one's own reasons and one's own values, when one wishes to. Impeding a person's ability to make such decisions and to incorporate her values and reasons into important decisions is morally problematic not because it may harm her (in many cases it will not), but because it undermines, or fails to respect, her autonomy. So, in the medical context, a person's decision whether she will undergo a procedure is an important one, and respecting her autonomy demands ensuring both that she has adequate information to determine whether the decision comports with her own reasons and values, and that she be able to decide even if she makes a choice that is harmful to her.

Accounts linking privacy and autonomy are highly varied, but there are three components important here. First, privacy may be an *object* of autonomous choice. People value privacy, in at least some respects, and may conceive of their goals, projects, and actions as being their own, and not for disclosure to others. Moreover, people may value acting and making decisions without others' observations. The fact that people value privacy—that is, it is a value that people hold autonomously—is a reason that others ought to respect that privacy (see Benn, 1984). In other words, privacy may be constitutive part of a life (or parts of a life) that people consider valuable.

A second facet is that privacy may be *condition* of autonomy. A number of commentators have argued that others' access to information about one's habits, activities, opinions, feelings, aspirations, and the like can undermine the degree to which one acts or thinks for oneself. Under scrutiny, one may implicitly incorporate others' values into one's beliefs and decisions, which in turn makes those beliefs and decisions less one's own (Bloustein, 1964; Reiman, 1976). Based on its relation to autonomy, a number of commentators have maintained that privacy is an important condition for liberal democracy (see de Bruin, 2010).

There is a third way in which privacy and autonomy connect. As noted, acting autonomously requires having sufficient information that one can make important decisions according to one's values and desires, and respecting autonomy demands that one disclose such information to others. However, information may also be important where it would have no bearing on one's actions. Instead, information may be important insofar as it allows people to interpret their situation in the world. Deceiving people, or limiting their access to important information, prevents people from seeing aspects of the world and limits their ability to interpret the world, regardless of whether they would act any differently (Hill, 1984).

Other views locate privacy's value in its relation to human flourishing. Adam Moore (2003, 2010) argues that where people do not have the ability to control others' access to them and their information, it undermines goods essential to human flourishing: mental and physical health, social relationships, and so forth. Julie Cohen links privacy to the *capability approach*, which takes the ability to exercise fundamental human capabilities as a primary moral good; threats to privacy, on this view, jeopardize people's abilities to develop those capabilities and flourish (Cohen, 2012: 225-229; see also Sen 1999; Nussbaum).

The range of views of privacy's nature and value suggests that privacy is valuable in different ways, and that we have to look at the specifics in order to determine whether any particular case (or policy, practice, etc.) is justified all things considered. This is the approach Nissenbaum takes. On her view, defying entrenched privacy norms is a prima facie violation of contextual integrity, but determining whether it is problematic all things considered requires looking at a range of goods (well-being, autonomy, justice, flourishing, and the like) and whether the practice impinges the values of the relevant context (Nissenbaum: 2010, 182-183).

3. Privacy and Learning Analytics

The professional, popular, and scholarly literature recognizes the potential conflict between learning analytics and student privacy, generally in the context of discussions about learning analytics more broadly. Our aim here is to contribute to that debate by focusing on privacy and, hence, providing a systematic treatment. There are in our view five key questions to ask in addressing privacy and learning analytics. That is, before it is reasonable to conclude that it is morally permissible to pursue learning analytics, or to pursue some particular form of learning analytics, we should address the following five problems. The first four we will refer to as the "narrow" questions. These are: (1) privacy and information flows *with respect to whom*; (2) privacy *about what*; (3) how to fully weigh the benefits and burdens of information collection; and (4) the extent to which various stakeholders, especially students, know about, choose, or endorse information collection, analysis, and use. We call these narrow questions not because they are easy to answer or of limited importance. Rather, they are narrow in the sense that they are consistent with the overall learning analytics enterprise and do not conflict with its basic premises. At most they demand policy changes about what and how information is collected and analyzed. In contrast, the fifth, "wide" question asks whether collecting information about students, learning environments, and outcomes conflicts with the values that justify higher education in the first place. As a result, it calls into question the project of surveillance in the service of higher education learning outcomes.

The Narrow Questions

1. Privacy with respect to whom.

In the previous section we explained that privacy must be understood as a three-part relation between some person (P), some domain of information (O), and some other person or persons (Q). In the context of learning analytics, discussions of student privacy are implicitly about privacy with respect to the institution. However, being specific about domains of information and other persons is crucial in understanding student privacy in learning analytics for several reasons.[1]

For one, information may be relevant to, and justifiably collected and analyzed for, one educational context, but not another. Suppose, for example, that it is useful to track students in courses with online components, and Z University decides to collect information about students watching presentations in a particular course's web page within a LMS. ZU collects information about which students view which presentations, when they view them (time of day, proximity to exams), how many times they view them, and so forth in order to determine whether there is a link between viewing practices and performance in the course. It is of course plausible that viewing practices correlate with course performance. However, even if such monitoring by the institution is justifiable in order to understand the connection between viewing habits and academic performance, it does not follow that the information should be available to (for example) course *instructors*. That is, it may be justifiable for ZU (Q1) to collect information about presentation viewing habits (O) of student P on the grounds that combining that information with lots of other information about other students and their viewing habits could lead to insights regarding behaviors and learning outcomes.

It also does not mean that students should have the information (here P and Q are the same). Information regarding viewing habits may be used to predict student success (or failure). And when failure is the probabilistic judgment in a red light, yellow light, green light intervention system such as Course Signals, the student may not necessarily want to see that information. In fact, such information may have adverse effects on her self-efficacy, such as a feeling of stigma (e.g., "I'm a red student; therefore, I'm no good") (Johnson, 2012).

Moreover, the possibility of gaining generalized knowledge about behaviors and learning outcomes does not provide a reason that student P's privacy regarding her presentation viewing habits (O) should diminish with respect to her *instructor* (Q2). Indeed, having access to information about how students are behaving in non-graded aspects of a course could unfairly prejudice an instructor for or against that student. Suppose, for example, that P waits until the day before the test to view all of the online materials. If the instructor sees this, she could grade P more harshly on the grounds that P did not prepare diligently ahead of time. Or, if the instructor sees that P views all presentations in a timely manner, and multiple times, she might give P the benefit of the doubt in P's graded work. Hence, where information in domain O is available to Q2—the instructor—it may conflict with the goal that students be evaluated based solely on the quality of their work.[2]

[1] As Nissenbaum argues, establishing key actors (senders and receivers of information, subjects of information) is key in analyzing moral claims (Nissenbaum, 2010, p. 149).

[2] Note, too, that instructors themselves may have reasons to not have access to information about students' activities within a course site in an LMS. Consider the case of a rigorous, technically demanding course aimed at students with a range of backgrounds, preparation, and comfort with the material. Some students may need to spend what seems to them an inordinate amount of time working through practice exercises, reviewing material, and rewatching presentations on the course site. In order to avoid the possibility that the students would be self-conscious and hence avoid spending lots of time preparing within the course site, an instructor may wish to

Let us be clear that we are *not* arguing that instructors should never have access to such information—far from it. Rather, the point is that the justifications for one party's access to a particular category of information will oftentimes be different from the justifications for other parties' access to that information. So, one might argue that instructors ought to be able to see information about students' activities on a course page within an LMS so that instructors can use the information to directly evaluate student performance or to troubleshoot for students doing poorly in the course. That is a plausible enough argument, but it is a *distinct* argument from the one that the broader institution can conduct research about correlations between behaviors and course performance and conduct early interventions. Moreover, the argument that the instructor should be able to use information collected within a course LMS page is contingent upon the instructor wanting to use information in that way and not using it in a way that would be unfair to the student (e.g., in such a way that could lead to unconscious or even conscious biases).

Another reason that we must be careful in considering the particular POQ relations is that the mere fact that information is collected in the context of a course for the benefit of the instructor does not justify its collection by the institution. Suppose, for example, that an instructor has students take weekly quizzes in a course page and has students upload essays into a course drop box. In that case, students' (P) privacy regarding their quiz performance and writing skills (O) diminishes with respect to the instructor (Q1). That is of course necessary for the instructor to evaluate the students, and generally based on the instructor's plan for the course. Information about student quiz scores or writing abilities might also be useful for the institution's (Q2) own purposes—perhaps evaluating the development of writing and quiz-taking skills over the course of students' time at the institution. Suppose, however, that the instructor wishes the students to take quizzes and write essays without any worries that information about them is being shared with a broader audience. It is at least plausible that the ability to make such choices within the university's LMS is consistent with instructors' freedom to teach courses in the manner they deem most conducive to learning.

A third reason to be careful about the entities that may access information via learning analytics is that there are plausible scenarios in which information gathered and analyzed in the service of learning analytics could be shared with third parties. The Family Educational Rights and Privacy Act (FERPA), protects student records from disclosure to third parties, except under certain exceptions. Student data collected via LMSs or other campus information systems appear to be "records" under FERPA (Family Educational Rights and Privacy, title 34, sec. 99.3); hence, it is legally protected, and universities will no doubt take appropriate measures to keep student information secure. However, it by no means follows that the information will remain private with respect to third parties. FERPA affords students the right to inspect their own records (Family Educational Rights and Privacy, title 34, sec. 99.10). And once students have access to those records, nothing prevents them from releasing the records to others, for example in order to secure a job, apply to a different educational institution, or to have a background check. If learning analytics lives up to its promise and becomes a useful tool to

prevent the LMS from either collecting the information or making the information visible to the instructor. Thanks to C.H. for this example from CH's own experiences.

predict academic (or other) success, it is hard to imagine that other third parties (e.g., insurance companies, creditors) would not seek to obtain information from learning analytics systems, especially potential employers. After all, employers already demand transcripts, letters of recommendation, and standardized test scores, and it is easy to see why they would value information tracking students' behaviors and work habits over time. And even though the records are protected under FERPA, fierce competition for jobs will ensure that many students will agree to provide the information. Perhaps this is a good thing. Richard Posner (1984), for example, argues that allowing individuals to conceal information about themselves is economically inefficient, and leads to overall decreases in social welfare. But the important point is that it is a mistake to treat instructors and institutions as the only relevant third parties in assessing privacy in learning analytics.

A related issue is that once created, student records may be subject to collection by yet another third party—government actors outside the education sphere. Specifically, records may be subject to collection pursuant to a warrant or subpoena. To use just one example, the section 215, "business records" provision of the USA Patriot Act has been at the center of a recent controversy regarding the U.S. National Security Agency's bulk collection of telephone metadata. Under section 215, the director of the FBI may "request" any "tangible thing"—including business records—that is relevant to foreign intelligence investigations, and education records such as those developed in learning analytics appear to fit (USA Patriot Act, sec. 215). Now, this example may seem far-fetched, but just as proponents of learning analytics want "everything," so too do other entities. And creating a whole new category of well-sorted and insightful information may elicit broader attention. One may argue that such availability is overall a good thing, or that merely collecting the information does not make one responsible for other uses to which that information might be put. However, if we are to fully capture the privacy implications of learning analytics, we should account for the fact (and it is indeed a fact) that educational institutions, who may well seek only to further learning objectives, are not the only Qs who matter.

We can draw several lessons here. First is that when considering the merits of learning analytics, we must do so with an eye on the particular POQ relation involved. Second, the particular POQ relation involved will implicate different values—making information available to the student might affect flourishing, making information available to an instructor might undermine the fairness of evaluation, and so forth. Third, there should be controls in any information system that allow for differential access. Thus, the decision to allow (e.g.) instructors, advisors, or others access to information collected from the LMS by the institution should be made deliberately and based on a consideration of the merits of that access. In the instructor case, it could be an option to collect and have access to the information, and the instructor would have to expressly choose such access based on her deliberate decision to use that information in evaluating students or gauging student engagement, or the like. And as we will discuss below, students should be able to know whether such information is collected, and who has access to it. Fourth, when we are explicit about the entities that might gain access to student information, lots of values outside the educational context come into play (e.g., employment, credit, law enforcement and security).

2. Privacy about what

The next question concerns the types of information that should be collected. We have seen that often the criterion proponents use in determining whether information should be collected analyzed is whether it is relevant to assessing learning environments and improving outcomes. Call this the “relevance view.” Notice, though, that because we cannot know *a priori* what information will shed light on learning environments and outcomes, any information would seem fair game on the relevance view. This is aligned with the view that in the context big data, data should be collected first and questions should be asked of it later; statistical sampling procedures do not apply (Mayer-Schönberger & Cukier, 2013). It is therefore not surprising to see proponents proposing to collect “any data [they] can get their hands on” or to take a “smorgasbord” approach to data gathering (Diaz & Brown, 2012, p. 13). This, however, assumes that the learning-related consequences are the *only* morally relevant consideration. But surely that’s not the case.

Suppose that ZU is interested in determining what factors lead to student success in certain courses and majors. It is plausible that students’ religious affiliations (Does P attend services regularly? Which services?), political activism (Is P active in campus, local, national, or international political activities? Of what sort and to what extent?), and social circles (Who are P’s friends and acquaintances? How academically successful are they?) are relevant in predicting academic success. Perhaps strong sectarian ties are correlated with high GPAs in some fields, regular attendance at religious services is associated with higher graduation rates, political activism is correlated with low class attendance and higher rates academic probation, and strong social ties to students with low grades is correlated with higher dropout rates. Perhaps instead the correlations are reversed. Regardless, if the *only* criteria for whether it is justifiable to collect information about students is its relevance to learning environments and learning outcomes, then it would follow that a university is justified in collecting information about students’ religious affiliations, political activities, and social networks. It is, however, not at all clear that universities (and especially state supported universities) ought to be in the business of collecting such information about their students.

Here we should nip three potential objections in the bud. First, one might argue that it is a university’s business to learn anything that is relevant to academic outcomes. Such an argument, however, would simply assume that the relevance view is correct. But the question here is *whether* relevance to academic outcomes is the only criterion that universities should consider in collecting information about students. To argue that religious, political, and social information could be collected on the grounds that it may be relevant would therefore be question-begging.[3]

[3] Even if one were to accept that universities may justifiably collect information in these domains, it is hard to imagine that there is literally *no* type of information that universities could not obtain on the grounds that it may be relevant to academic success: sexual orientation? Sexual habits? Mood? Weight gain? Fashion sense? Diary entries? The key point here is that the relevance view admits of *no* functional limits on information collection because no information can be determined irrelevant *a priori*. Whether the information collected is justifiable on relevance grounds can only be determined once the information has been collected and analyzed.

Second, one might argue that such factors are not *causally* related to learning behaviors and outcomes, and that they would more likely be evidence of some underlying factor that affects learning. Perhaps, but if they are good evidence of those underlying factors, they would indeed seem to be relevant to the goals of learning analytics and hence fair game on the relevance view. Third, one might object that universities have no mechanism to collect such information, so there is little worry that they will begin monitoring students in such problematic domains. However, insofar as other information systems (geolocation, social media) are attractive sources for data (as noted above) there is no reason to rule out the possibility that more sensitive information could be collected in the context of learning analytics.

Regardless, the broader point is not about these particular types of information. Rather, it is that there are at least *some* types of information that ought not be collected in an attempt to reach better learning environments and educational outcomes. Mere relevance is not enough. But if that is correct (and denying that it is correct entails that collecting religious, political, and associational information would be permissible), we need some set of criteria for determining whether it is permissible to collect information.

Those criteria would be subject to debate, but a starting point might be that any information about student behaviors that universities have no legitimate reason to influence is impermissible to collect. Recall that one of the key premises of learning analytics is that information about student behaviors can be analyzed and universities may intervene in cases where students perform poorly or are at risk of not maintaining good academic standing. Hence, a good principle for whether it is justifiable to collect information is whether the universities would be justified in intervening to change the behavior captured in the data. Suppose (to use an example from above) it turns out that political activism is correlated with high dropout rates or low grades. If some student was both getting low grades and involved in politics, and her university had information about both (say, from a LMS and from social media information collected), it could conceivably intervene by recommending that the student spend less time on her activism so as to remain in good academic standing. But if it is illegitimate for a school to intervene in that way (i.e., by making a suggestion about her political activities), then collection of information about those activities is unjustified. That is, the action that would derive from collecting the information is itself illegitimate, which undercuts the purpose of the information collection itself.

But why would it be illegitimate for a university to intervene about a student's political activities? Here we can draw on the idea of liberal neutrality. The idea, roughly, is that in a liberal democracy, state policies and practices ought not favor or disfavor any particular conception of the good; that is, state actors are constrained in the ways in which they can help advance or hinder different ideas of the good life. (Raz 1986: 110; Rawls 1996: 191-195; Nozick 1974: 272-273). There are a number of ways to interpret political neutrality. One is that state action may not be *grounded* on favoring some views over others. Alternatively, it might mean that no state action may affect the likelihood of a person endorsing or realizing her conception of the good. (Raz 1986: 114-115). In the case of learning analytics, the purpose of collecting and acting on information is not grounded in favoring some conceptions of the good (or, at least, it favors only conceptions that are endorsed by those affected—people pursuing an education). But even if this is the *kind* of reason on which a state may generally base policies, it operates directly on persons' pursuits of important facets of their lives and on their particular conceptions of the good. Put another way: things like religious observance (or lack thereof), political activity (or lack thereof), and interaction with people of one's choosing are central to many people's conceptions of the good. Seeking to affect those types of activities would for some people substantially hinder their ability to realize their conceptions of the good, contravening the second interpretation of liberal neutrality.

Two objections are worth addressing. One is that we have no reason to think that any school (much less a state school) is considering intervening on these bases. But, as we argue above, the premise of collecting the information in the first place is affecting student behaviors and learning outcomes; but if interventions based on politics, religious observances, and social circles are impermissible, then the information collection is itself unjustifiable. So, the mere fact that there are no plans to use the information is neither here nor there. Another potential objection is that our argument is applicable only to state universities, and hence not an issue for non-state institutions. That may be true strictly as a matter of liberal political theory. However, to the extent that private institutions seek to respect and promote reasonable pluralism (and hence not favor particular conceptions of the good), they would be under similar constraints by virtue of the values they endorse.

3. Proper Accounting Of Benefits And Burdens

Proponents of learning analytics maintain that collecting, analyzing, and using information from LMSs, SISs, and other sources will be beneficial (Campbell, DeBlois, & Oblinger, 2007; Mayer-Schönberger & Cukier, 2014). If we learn more about student behaviors and learning outcomes, the argument runs, institutions will be better able to tailor teaching, intervene where students are at risk, and allocate resources in ways that are conducive to educational goals. At this early stage it is hard to argue against the possibility of benefits. Nonetheless, it is an open question as to whether the possibility of good consequences overall is sufficient to warrant the sort of information collection involved in learning analytics. There are several concerns here. One is whether consequences alone can justify the privacy losses inherent in learning analytics. Another is that institutional interests may diverge from student interests. A third is that the distribution of benefits among students will vary, potentially unfairly.

Rights and consequences.

The primary reason in favor of aggressively pursuing learning analytics is that there will be benefits, and that those benefits outweigh potential costs. Implicit in this formulation is that whether or not learning analytics is justified turns solely on its consequences, and the reasons we have for pursuing, or not pursuing, learning analytics begin and end with the consequences that will result. It is, however, entirely unclear whether we should simply weigh the consequences of learning analytics in order to determine its value. There are after all lots of circumstances in which the beneficial consequences of actions are not legitimate considerations at all. Consider promise-keeping. Suppose, for example, that Dora has promised to water Gina's plants while Gina is out of town. But let's say that the weather is cold and rainy, and it would be very unpleasant for Dora to make the trek to Gina's place and water the plants. They'll probably survive anyway, at most looking a bit ragged with Gina returns. Dora's discomfort is not only an insufficient reason to break her promise to Gina, it is not even a relevant factor. The promise is an "exclusionary" reason, as Joseph Raz puts it (Raz, 1999, p. 37; see also Waldron, 2003, p. 196).

Consequences may also be irrelevant where there are rights involved. So, for example, Robert Nozick understands rights as “side-constraints,” which prohibit outright certain actions regardless of consequences, and Ronald Dworkin understands rights as moral “trumps” which operate independently of the consequences they engender (Nozick, 1974; Dworkin, 1984).

The question here is whether privacy operates as a right, or as an exclusionary reason. Certainly privacy operates as such a reason in some circumstances. In the law enforcement context, constitutional and statutory protections for privacy (e.g., rights against unreasonable searches and seizures, or statutes limiting the ability of police to conduct electronic surveillance) operate even if those constraints hinder law enforcement. Above we describe accounts that place value on privacy as deriving from, or necessary for, individual autonomy. These contrast with consequence-based arguments for privacy’s value. Hence, respecting privacy may be important even if good consequences from surveillance would go unrealized. As such it may be protected as a matter of right, and hence it would take more than marginally better consequences to infringe.

Allocation of benefits

Even if we suppose that the consequences of learning analytics are overall positive, and that privacy ought not be protected by a right, there is a question as to who it benefits, and to what degree. It is certainly the case that learning analytics will be a benefit to institutions. And to the extent that institutions will use the information to further their mission of providing learning opportunities and helping ensure learning outcomes, some benefits would presumably accrue to students as well. But that does not mean that the distribution of benefits is good, or fair.

To begin, institutional benefit and student benefit may diverge; institutional goals and student benefits may also align (see, for example, Long and Siemens (2011) list of aims for which learning analytics may be put in section 2). But institutional goals and student benefits are not identical, and conflating them risks subordinating student benefits to institutional goals. One possibility is that knowledge derived from tracking students will be used to improve instruction and the overall educational environment. However, institutions wish to improve student performance, retention, and graduation. One way to do this is by improving instruction. But it may be far easier to use information to change recruiting practices or to steer students toward classes in which they are more likely to do better. Those measures would not clearly be in the interests of students overall (and is clearly *not* in the interests of students who would otherwise be admitted), even though they would plausibly be in the interest of the institution.

In addition, the students that will benefit are not identical to those whose information is collected. The hope is that learning analytics will use data about individual, identifiable students to make beneficial changes in their academic lives. Notice, though, there is a difference between benefiting students overall, and benefiting the students whose information is collected. In order to understand data collected and its relation to learning, researchers will have to study students who will not benefit from the information collection—their educations will be complete by the time conclusions become actionable. Moreover, even if we suppose that learning analytics will improve learning environments and outcomes, it is an open question how those benefits will be distributed.

Consider a different example about the distribution of benefits. Suppose that an entry level statistics course in an engineering program is graded on a strict curve whereby 25 percent of students will receive too low a grade to continue in the program. In the first half of the course, the learning analytics system identifies the students likely not to pass, and the instructor takes measures to provide them with extra help. If these students succeed, other students will *necessarily* do worse.[4] The system distributes a benefit to some, and imposes a cost on other students, though the institution benefits insofar as students overall learn more. Perhaps this is a good thing, perhaps not. Regardless, there is a question of what a just distribution of consequences is in this case, and (hence) we cannot say that the analytics involved are justifiable just because there is some benefit from the standpoint of the institution.

One final consideration is of paramount importance here, namely, the counterfactual case. Substantial resources are going toward learning analytics, and as Slade and Prinsloo (2013) point out, expense is a key concern higher education administrators have regarding learning analytics, citing an uncertainty as to whether learning analytics just a limited "endeavor" or long-term "investment" in the future (Bichsel, 2012, p. 3). It is entirely unclear, though, whether the resources spent on data analytics will lead to as much educational benefit as other possibilities: increasing faculty-student ratio, increasing academic support services, providing more resources to primary and secondary education, or something else altogether.

4. Awareness and Control

A fourth question concerns the relationship between students—the subjects of learning analytics—and the practices surrounding learning analytics. Regardless of who has access to information about students within some domain, the justifiable scope of information collected in the context of learning analytics, and how the benefits of learning analytics are distributed, there is an issue regarding students' involvement in the process of collecting and analyzing their information. Specifically, should students be aware that, and the degree to which, their information is collected and analyzed? And should students have meaningful choice as to whether, and to what degree, their information is collected and analyzed.

As learning analytics is still an emerging technology and practice, and because it is used primarily by instructors and administrators, students have little awareness about its purpose, goals, and the data it collects. And institutions may have good reason to not raise student awareness of learning analytics practices. For example, institutions may fear that if students become aware of the depth and breadth of data they use about their student body that it may negative repercussions: protests, legal action, avoiding institutions, courses, or activities that collect information. And student awareness may create a chilling effect on the student body.

It may not always be the case, however, that student awareness has detrimental effects. Even so, the fact (if it is a fact) that students are unaware of monitoring is important with respect to their autonomy interests. As noted above, an important element of respect for autonomy is provision of

[4] Again, thanks to C.H. for an example from C.H.'s experience.

information important for people to understanding their situation. Whether one is monitored within an LMS would seem to be important information. This creates an impetus for institutions to make students aware of data practices, benefits, potential concerns, and the like. In doing so, the institution would create a sense of transparency about learning analytics and develop trust among the student body. Furthermore, instructors could include syllabus statements regarding their use of learning analytics, the ends to which it will be put (e.g., assessment, purposeful interventions, advising), and the advantages of using learning analytics in the classroom.

Student Choice

A related question is whether students should have some say in whether, and to what extent, their information will be used. In section 3 we saw that privacy may be important as an object of autonomous choice. People value privacy (in at least some cases and to some degree), and as Benn (1971) argues, where people value their projects, actions, and choices in part because they are able to do them without observation, we have a reason to afford them that opportunity out of respect for their ability to determine what is important to them. In turn, respect for students implies that they should have ample opportunity to opt out of information collection. It is not enough that they could attend other institutions, or opt out of higher education altogether. The ability to choose among unattractive options does not make for an autonomous choice (see Raz, 1988, pp. 408-410).

This notion of choice finds support in the Family Educational Rights and Privacy Act (FERPA) of 1974, which sets guidelines on institutional management of educational records. As noted above, it appears that learning analytics data is a part of one's educational record. FERPA provides a number of baseline protections for student privacy, including a right of inspection, a right to amend inaccurate or misleading information, and a right to file a complaint where one's rights are violated. A student has no statutory right to prevent use of her educational records by institutional actors and school officials with a "legitimate educational interest" in the information. However, institutions could provide students with choice over their data. The Department of Education often refers to FERPA as the "floor" for protecting privacy and not the "ceiling" (Family Policy Compliance Office, 2011, p. 5). Institutions could adopt opt-in or opt-out processes for learning analytics. Or, to provide greater control to students, institutions could create a data management dashboard for students which would enable them to turn on and off particular sources of data, data types, and information about themselves for use in learning analytics practices or by the institution as a whole. Again, insofar as autonomy is an important facet of privacy, student choice is how their data is treated is important for respecting values that underwrite privacy protections.

Summing up, the narrow questions draw on the fact that privacy is multifaceted and implicates a range of values in a variety of ways. Analyzing it demands being specific about particular persons or groups of persons, domains of information, and other parties. Loss of privacy may affect well-being, it may undermine autonomy, and the fact (if it is a fact) that benefits accrue from information gathering does not by itself justify information collection. Information gathering may reinforce already-existing social advantages, it may be used to benefit other people than those whose information is gathered, and the results of interventions may be distributed unfairly. Moreover, privacy implicates what we might call second-order values: the fact of privacy loss may not be as important in some cases as being aware or unaware of the loss, or the ability to decide whether to participate or endorse information collection. That range of issues isn't captured by reducing privacy to a univocal value.

The Wide Question

The four narrow questions are tractable. It is possible for different parties to have varying degrees of access to student information in different domains, there could be limits on the types information collected for the purposes of learning analytics, we could work to ensure that learning analytics tends to advance student, rather than institutional, interests and that it does so fairly, and we can ensure that students are aware of information collection and have reasonable options to avoid it. The wide question, however, is more difficult. Recall from section 3 that several important justifications for privacy protections are based on persons' autonomy. As we've seen, a number of commentators argue that privacy is necessary for one's values, choices, and projects to be one's own, and to have "moral title" to one's actions. Part of the reason autonomy is important is that it requires a degree of respect for individuals; to the extent people have their own reasons and values, and can act according to those reasons and values, we have a responsibility to afford them the ability to so-act. Another reason that autonomy is important is that it is a condition for democratic legitimacy. One tenet of liberal democratic theory is that legitimate government must be based on consent of those governed, and autonomy is a necessary condition for consent (see, e.g., Brighouse 1998). To the extent that privacy is a condition of autonomy, it underwrites that legitimacy.

But what does this have to do with information collection in higher education and the growth of learning analytics? The justification for the information collection and analysis at the heart of learning analytics is better learning outcomes. However, in order to understand why better learning outcomes are desirable at all requires that we consider what justifications there are for higher education in the first place. After all, the outcomes measured in learning analytics will include (for example) grades, retention, graduation rates, and the like. But those are not the ultimate goals of higher education; if they were, universities could simply give higher grades and diplomas to all matriculated students. Such measures are instead (hopefully) evidence of some other, further goods that underwrite higher education.

In *Our Underachieving Colleges*, Derek Bok canvasses a number of views of the purposes of higher education and rejects the idea that there is a single, unifying purpose of higher education. Rather, he argues that there are a variety of values and goods that should guide higher education:

> [A]ttempts to prescribe a single overriding aim or to limit the purposes of college to the realm of intellectual development take too narrow a view of the undergraduate experience and threaten to impose a moratorium on efforts to nurture some extremely important human qualities during four formative years of students' lives. Instead colleges should pursue a variety of purposes, including a carefully circumscribed effort to foster generally accepted values and behaviors, such as honesty and racial tolerance. Within this ample mandate, several aims seem especially important (Bok, 2006, p. 66).

These aims include communication, critical thinking, citizenship, living with diversity, living in a more global society, and employment (Bok, 2006, pp. 67-81). This is a certainly a plausible list, and any account of the value of higher education should be able to explain its relation to each element (see, for example, Gutmann, 1980, pp. 200-201).

Notice two things. First, these elements are not directly measurable, with the possible exception of "preparing for work," which one might measure by looking at employment and income data. The fact that one of these values is more amenable to data collection and measurement is troubling insofar as it may take on a disproportionately large role in determining what counts as "success" in learning outcomes. That may well create incentives for universities to "nudge" students toward majors and courses in which they are more likely to succeed (regardless of whether that is the student's actual interest) and to tailor curricula to find more success on this one measure. That is not to say that work and careers and unimportant. They are. Rather, the fact (if it is) that they are important *and measurable* could lead them to take an even more prominent place in institutional goals.

More important, though, is that at least *some* of the values described by Bok are partially constituted by an ability to exercise autonomy, that is, to act according to one's own reasons and values as one sees fit. Consider citizenship. John Rawls's view of a just, democratic state is based on a conception of citizens who are free and equal, and who are reasonable and rational (Rawls, 2001, pp. 6-7). They are reasonable in that they can abide fair terms of cooperation, so long as others do as well. They are rational insofar as they can formulate, revise, and pursue their own sense of what is valuable. Both of these elements of citizenship are closely aligned with autonomy, or the ability to act according one's own reasons as one sees fit. Other items on Bok's list are *component* parts of autonomy. Take critical thinking. As Harry Brighouse notes in the context of education, "broadly speaking, the capacities involved in critical reflection help us to live autonomously" (Brighouse, 1998, p. 728). Likewise with effective communication, living with diversity, and living in a diverse world.

Returning to information gathering and privacy, if it is indeed the case that privacy is an object of autonomy (something people, including students, value for their own reasons) and a condition of autonomy (ensuring that people act for reasons that are their own), the consistent, wide-ranging information collection that is a central feature of learning analytics poses at least *some* conflict with students' autonomy interests. And to the extent that the purposes of higher education—that is, the values that underwrite it—include goods that are based on autonomy, then the collection, analysis, and use of large swaths of student information to *measure* learning outcomes conflict (potentially, and to some extent) with the goods that the measures should support. In other words, diminishing privacy to advance higher education's aims may serve instead to undermine those aims.

There are important rejoinders here. One might doubt the link between privacy and autonomy. Perhaps it is not true that being monitored affects people's actions, and perhaps people do not internalize the (perceived) values of others and act differently and for reasons that are not their own due to monitoring. That is both a conceptual question (what does it take for one's actions to be "one's own" in the right sense?) and an empirical question (how much does monitoring actually affect behavior?). If it turns out that people do not act for reasons that are not their own in the right sense, then the wide problem is not a problem.[5]

Another is that if learning analytics enables students to do better at whatever course of study they choose, and which instantiates the purposes of higher education, then they will *overall* be better able to act according to their own values and reasons. Again, this may be true, but it is contingent on how learning analytics is actually developed, implemented, and used.

In the end, the important aspect of the wide question is that it suggests at least one criterion for whether a learning analytics system is justifiable, viz., whether it promotes or conflicts with persons' autonomy. To the extent that it can aid institutions in helping students learn to communicate, think critically, and live with diversity in a more global society, it promotes their autonomy. And the information collection does not undermine the degree to which students actions are their own, and done for their own reasons and values, it is consistent with their autonomy interests and justifiable. On the other hand, if learning analytics tends to push students in directions are not based on their own reasons, or promotes some values (e.g., careers) disproportionately to others, it conflicts with their autonomy interests and is not justifiable.

5. Conclusion

Learning analytics presents significant student privacy problems for higher education institutions. We have argued that before we conclude that learning analytics is justified, proponents must address four narrow problems related to the use of student data, and we have posited partial answers to each: 1) learning analytics systems should provide controls for differential access to private student data 2) institutions must be able to justify their data collection using specific criteria–relevance is not enough; 3) the actual or perceived positive consequences of learning analytics may not be equally beneficial for all students, and the cost, then, of invading one student's privacy may be more or less harmful, and we need a full accounting of how benefits are distributed between institutions and students, and among students; finally, 4) in spite of legal guidelines that do not require institutions to extend students control their own privacy, they should be made aware of collection and use of their data and permitted reasonable choices regarding collection and use of that data. We also argued that there is a wider question concerning learning analytics: the practice may diminish student privacy to the detriment of a student's autonomy, which is related to some important values underwriting higher education. To the extent that higher education is important as a function of autonomy, learning analytics is justifiable just to the extent that it does indeed promote autonomy.

References

Allen, A. L. 1988. *Uneasy access: Privacy for women in a free society*. Totowa, N.J.: Rowman & Littlefield.

[5] Though if we were to doubt the connection between privacy and persons' acting according to their own values as they see fit, we might also have to forego objections to lots of other types of information gathering and surveillance.

Allen, I. E., and J. Seaman. 2013. *Changing course: Ten years of tracking online education in the United States.* Survey Report. http://sloanconsortium.org/publications/survey/changing_course_2012 (accessed April 18, 2014).

Alter, A. 2012. Your e-book is reading you. *The Wall Street Journal*, July 19, 2012. http://online.wsj.com/news/articles/SB10001424052702304870304577490950051438304 (accessed April 18, 2014).

Arnold, K. 2010. Signals: Applying academic analytics. *EDUCAUSE Review*. http://www.educause.edu/ero/article/signals-applying-academic-analytics (accessed April 18, 2014).

ASU. 2011. New initiative advances ASU's effort to enhance student success. *ASU News*, October 12, 2011. https://asunews.asu.edu/20111012_eAdvisor_expansion (accessed April 18, 2014).

Barber, R., and M. Sharkey. 2012. Course correction: Using analytics to predict course success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge,* 259–262. Vancouver.

Benn, S. I. 1971. Privacy, freedom, and respect for persons. In *NOMOS XIII: Privacy*, edited by J. Roland Pennock and John W. Chapman, 1–26. New York: Atherton Press.

Bichsel, J. 2012. Analytics in higher education: Benefits, barriers, progress, and recommendations. ECAR Report. http://net.educause.edu/ir/library/pdf/ERS1207/ers1207.pdf (accessed April 18, 2014).

Blaauw, M. 2013. The epistemic account of privacy. *Episteme* 10 (Special Issue 02): 167–77. doi:10.1017/epi.2013.12.

Bloustein, E. 1964. Privacy as an aspect of human dignity: An answer to Dean Prosser. *New York University Law Review* 39: 962–1007.

Bok, D. C. 2006. *Our underachieving colleges: A candid look at how much students learn and why they should be learning More*. Princeton, N.J.: Princeton University Press.

Brown, M. 2011. *Learning analytics: The coming third wave*. ELI Report ELIB1101. http://net.educause.edu/ir/library/pdf/ELIB1101.pdf (accessed April 18, 2014).

Bruin, B. 2010. The liberal value of privacy. *Law and Philosophy* 29 (5): 505–34. doi:10.1007/s10982-010-9067-9.

Carmean, C., and P. Mizzi. 2010. The case for nudge analytics. *EDUCAUSE Quarterly* 33(4): n.p. http://www.educause.edu/eq (accessed April 18, 2014).

Chopra, A., and Z. Smith. 2012. Unlocking the power of education data for all Americans. *Office of Science and Technology Policy*, January 19, 2012. http://www.whitehouse.gov/blog/2012/01/19/unlocking-power-education-data-all-americans (accessed April 18, 2014).

Campbell, J. P., P. B. DeBlois, and D. G. Oblinger. (2007). Academic analytics: A new tool for all. *EDUCAUSE Review* 42(4): 40-57.

Cite redacted. In progress. Student privacy problems in a learning analytics infrastructure: A case study of a socio-technical reverse salient.

Cohen, J.E. 2012. *Configuring the networked self: law, code, and the play of everyday practice*. New Haven [Conn.]: Yale University Press.

DeCew, J. W. 1997. *In pursuit of privacy: Law, ethics, and the rise of technology*. Ithaca, N.Y.: Cornell University Press.

Denley, T. 2012. Advising by algorithm. *The New York Times*, July 17, 2012. http://www.nytimes.com/interactive/2012/07/18/education/edlife/student-advising-by-algorithm.html (accessed April 18, 2014).

Dennis, A., T. Duffy, and A. Morrone. 2010. *Indiana University e-textbook project*. Report. http://etexts.iu.edu/files/eTextbook%20Report%20-%20Spring%202010.pdf (accessed April 18, 2014).

Diaz, V., and M. Brown. 2012. *Learning analytics: A report from the ELI focus session.* ELI Report ELI3027. http://www.educause.edu/eli/ (accessed April 18, 2014).

EDUCAUSE Learning Initiative. 2011. *7 things you should know about learning analytics.* ELI Report ELI7059. http://www.educause.edu/eli (accessed April 18, 2014).

Family Educational Rights and Privacy Act. 1974. *Code of federal regulations*. Title 34. Department of Education.

Family Policy Compliance Office. 2011. *The Family Educational Rights Privacy Act: Guidance for reasonable methods and written agreements.* Document. http://www2.ed.gov/policy/gen/guid/fpco/pdf/reasonablemtd_agreement.pdf (accessed April 18, 2014).

Fried, C. 1970. *An Anatomy of values: Problems of personal and social Choice*. Cambridge, UK: Harvard University Press.

Grantees. n.d. *Next Generation Learning Challenges*. http://nextgenlearning.org/grantees (accessed April 18, 2014).

Grush, M. 2011. Monitoring the PACE of student learning: Analytics at Rio Salado College. *Campus Technology*, December 14, 2011. http://campustechnology.com/Articles/2011/12/14/Monitoring-the-PACE-of-Student-Learning-Analytics-at-Rio-Salado-College.aspx (accessed April 18, 2014).

Gutmann, A. 1980. Children, paternalism, and education: A liberal argument. *Philosophy & Public Affairs* 9 (4): 338–58.

Hill Jr., T. 1984. Autonomy and benevolent lies. *Journal of Value Inquiry* 18: 251–97.

Hoover, E. 2012. Facebook meets predictive analytics. *The Chronicle of Higher Education*, November 6, 2012. http://chronicle.com/blogs/headcount/facebook-meets-predictive-analytics/32770 (accessed April 18, 2014).

Ice, P., S. Díaz, K. Swan, M. Burgess, M. Sharkey, J. Sherrill, D. R. Huston, and H. Okimoto, H. 2012. The PAR framework proof of concept: Initial findings from a multi-institutional analysis of federated postsecondary data. *Journal of Asynchronous Learning Networks* 16(3): 63–86.

Jerome, J. W. 2013. Buying and selling privacy: Big data's different burdens and benefits. *Stanford Law Review Onlin*e 66: 47–53.

Johnson, J. A. 2012. Ethics of data mining and predictive analytics in higher education. Paper presented at the *Rocky Mountain Association for Institutional Research Conference*. Laramie, Wyoming.

Johnson, L., A. Levine, R. Smith, and S. Stone. 2010. *The 2010 horizon report*. Austin, TX: The New Media Consortium.

Long, P., and G. Siemens. 2011. Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*. https://net.educause.edu/ir/library/pdf/ERM1151.pdf (accessed April 18, 2014).

Mandge, O. L. 2013. A data mining tool for prediction of suicides among students. In *Proceedings of National Conference on New Horizons in IT,* 178–181. Rome.

Mattingly, K. D., M.C. Rice, and Z. L. Berge. 2012. Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal* 4(3): 236–247.

Mayer-Schönberger, V., and K. Cukier. 2013. *Big data*. Boston, MA: Mariner Books.

Mayer-Schönberger, V., and K. Cukier. 2014. *Learning with big data: The future of education*. Boston, MA: Mariner Books.

MyData. n.d. *Office of Educational Technology*. http://www.ed.gov/edblogs/technology/mydata/ (accessed April 18, 2014).

Nelson, L. A. 2013. Idea whose time has come? *Inside Higher Ed*, May 13, 2013. http://www.insidehighered.com/news/2013/05/13/political-winds-shift-federal-unit-records-database-how-much (accessed April 18, 2014).

Nissenbaum, H.N.. 2010. *Privacy in context: technology, policy, and the integrity of social life*. Stanford, Calif: Stanford Law Books.

Norris, D. M. 2011. *7 things you should know about first-generation learning analytics.* ELI Report ELI7079. https://net.educause.edu/ir/library/pdf/ELI7079.pdf

Norris, D., L. Baer, J. Leonard, L. Pugliese, and P. Lefrere. 2008. Action analytics: Measuring and improving performance that matters in higher education. *EDUCAUSE Review* 43(1): 42–67.

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Oblinger, D. 2012. “No more excuses”: Michael M. Crow on analytics. *EDUCAUSE Review*. https://net.educause.edu/ir/library/pdf/ERM1241P.pdf (accessed April 18, 2014).

Office of Science and Technology Policy. 2012. *Fact sheet: Unlocking the power of education data for all Americans.* Press release. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ (accessed April 18, 2014).

Parent, W.A. 1983. Privacy, morality, and the law. *Philosophy and Public Affairs* 12 (4): 269–88.

Parry, M. 2011. Colleges mine data to tailor students’ experience. *The Chronicle of Higher Education*, December 11, 2011. https://chronicle.com/article/A-Moneyball-Approach-to/130062/ (accessed April 18, 2014).

Parry, M. 2012. Big data on campus. *The New York Times,* July 18, 2012. http://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html (accessed April 18, 2014).

Posner, R. A. 1984. An economic theory of privacy” In *Philosophical Dimensions of Privacy*, edited by F. Schoeman, 333–45. Cambridge, UK: Cambridge University Press.

Rachels, J. 1975. Why privacy is important. *Philosophy and Public Affairs* 4 (4): 323–33.

Rawls, J. 1996. *Political liberalism*. The John Dewey Essays in Philosophy, no. 4. New York: Columbia University Press.

———. 1999. *A theory of justice*. Cambridge, M.A.: Belknap Press of Harvard University Press.

Rawls, J, and E. Kelly. 2001. *Justice as fairness: A restatement*. Cambridge, M.A..: Belknap Press of Harvard University Press.

Raz, J. 1988. *The morality of freedom*. Oxford, UK: Oxford University Press.

Reiman, J. H. 1976. Privacy, intimacy, and personhood. *Philosophy and Public Affairs* 6 (1): 26–44.

Richards, N. M., and J. H. King. 2013. Three paradoxes of big data. *Stanford Law Review Online* 66: 41–46.

Rubel, A. 2011. The particularized judgment account of privacy. *Res Publica* 17 (July): 275–90.

Rubel, A., and R. Biava. 2014. A framework for analyzing and comparing privacy states. *Journal of the Association for Information Science and Technology*, June, n/a – n/a. doi:10.1002/asi.23138.

Siemens, G. 2013. Learning analytics: The emergence of a discipline. *American Behavioral Scientist* 57(10): 1380–1400.

Solove, D. 2006. A taxonomy of privacy. *University of Pennsylvania Law Review* 154(3): 477–560.

Stanley, J. 2012. The potential chilling effects of big data. *American Civil Liberties Union,* April 30, 2012. https://www.aclu.org/blog/technology-and-liberty/potential-chilling-effects-big-data (accessed April 18, 2014).

Tene, O., and J. Polonetsky. 2013. Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property* 11(5): 239–273.

van Barneveld, A., K. E. Arnold, and J. P. Campbell. 2012. *Analytics in higher education: Establishing a common language*. ELI White Paper ELI3026. https://net.educause.edu/ir/library/pdf/ELI3026.pdf (accessed April 18, 2014).

Wagner, E., and P. Ice. 2012. Data changes everything: Delivering on the promise of learning analytics in higher education. *EDUCAUSE Review*. http://www.educause.edu/ero/article/data-changes-everything-delivering-promise-learning-analytics-higher-education (accessed April 18, 2014).

Wave I. n.d. *Next Generation Learning Challenges*. http://nextgenlearning.org/wave-i (accessed April 18, 2014).