Title

"We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education

Authors

Kyle M. L. Jones (corresponding author)
kmlj@iupui.edu
(317) 278-0046
Indiana University-Indianapolis (IUPUI)
School of Informatics and Computing
535 W. Michigan Street
Indianapolis, IN 46202

Andrew Asher
asherand@indiana.edu
(812) 855-1609
Indiana University-Bloomington
1320 E. Tenth Street
Bloomington, IN 47405

Abigail Goben
agoben@uic.edu
(312) 996-8292
University of Illinois at Chicago
Library of the Health Sciences
1750 W Polk Street (MC 763)
Chicago, IL 60612

Michael R. Perry
michael.perry@northwestern.edu
(847) 467-5488
Northwestern University
1970 Campus Drive
Evanston, IL 60208

Dorothea Salo
salo@wisc.edu
(608) 263-2900
University of Wisconsin-Madison
Information School
600 N. Park St.
Madison, WI 53706

Kristin A. Briney
kabriney@caltech.edu
California Institute of Technology

Mail Code 1-43
1200 E California Blvd
Pasadena, CA 91125-4300

M. Brooke Robertshaw
brooke.robertshaw@oregonstate.edu
(541) 737-1780
Oregon State University
1500 SW Jefferson Way
Corvallis OR 97331

Abstract

Higher education institutions are continuing to develop their capacity for learning analytics (LA), which is a sociotechnical data mining and analytic practice. Institutions rarely inform their students about LA practices and there exist significant privacy concerns. Without a clear student voice in the design of LA, institutions put themselves in an ethical grey area. To help fill this gap in practice and add to the growing literature on students' privacy perspectives, this study reports findings from over 100 interviews with undergraduate students at eight United States higher-education institutions. Findings demonstrate that students lacked awareness of educational data mining and analytic practices, as well as the data on which they rely. Students see potential in LA, but they presented nuanced arguments about when and with whom data should be shared; they also expressed why informed consent was valuable and necessary. The study uncovered perspectives on institutional trust that were heretofore unknown, as well as what actions might violate that trust. Institutions must balance their desire to implement LA with their obligation to educate students about their analytic practices and treat them as partners in the design of analytic strategies reliant on student data in order to protect their intellectual privacy.

Keywords: Higher education, student privacy, learning analytics, ethics, data

Introduction

Colleges and universities have adopted networked information and communication technologies at an unprecedented rate, enmeshing their campuses in an invisible layer of interconnected ubiquitous systems that regularly collect and store data about their communities. These systems serve needs such as administering complex institutions, supporting interpersonal communications and scheduling, facilitating face-to-face and online learning experiences, resource provision, and campus security, among other things. Whenever faculty, staff, and students gain access to, click on resources within, and submit information to these systems, their actions and content become data; and since institutional systems often require credentials for access, many of these data identify individuals (Jones, 2019b).

Academic institutions have begun to aggregate data to maximize data-driven analytic possibilities (Lane & Finsel, 2014). From these data they seek insights into specific behaviors and outcomes, craft interventions, and use strategies and algorithms adopted from data science, nominally for educational and administrative purposes. These strategies are commonly referred to as "learning analytics" (LA). Defined as a sociotechnical practice, LA is the "measurement, collection, analysis, and reporting of [student and other data] for the purposes of understanding and optimising learning and the environments in which it occurs" (Siemens, 2012, p. 4). Institutions claim that LA prepares them to describe (what is happening?), diagnose (why did it happen?), and predict (what is likely to happen?) factors that influence, enhance, or impede student learning, as well as prescribe (what should we do about it?) data-based interventions (Minelli, Chambers, & Dhiraj, 2013). In years past, there has been a clear distinction between so-called "academic analytics" and LA due to different (but not entirely dissociated) end goals, but over time the latter term has found more traction in the literature and everyday use (see Bienkowski, Feng, & Means, 2012; Cooper, 2012; Nunn, Avella, Kanai, & Kebritchi, 2016; Wong, 2017).

As is the case with most Big Data practices, higher education faces its own unique set of moral and ethical questions around LA, primarily student privacy. Student privacy is not a new concern for LA (Heath, 2014; Hoel & Chen, 2016; Ifenthaler & Schumacher, 2016; Jones, 2019b; Pardo & Siemens, 2014; Prinsloo & Slade, 2015), and the literature has evolved to address multiple facets of student privacy, but one major gap exists: the student voice has not been comprehensively addressed, especially in comparison with the overall advocacy for LA. Only lately has literature begun to address student perceptions of and expectations in regard to LA, especially where privacy is concerned. This gap in the literature needs rigorous research to infuse the student voice into technological designs, policies, and practices associated with LA. Our article aids in filling this gap.

To begin, we argue what student privacy is good for from a theoretical perspective. Next, we address why student privacy has become a paramount concern given new flows of data and information in support of LA and other educational data mining practices. This is followed with an overview of the literature that exists on student perceptions of LA regarding privacy. After describing our qualitative methods, we detail findings that demonstrate the lack of knowledge students have about LA, what students think might be possible with LA, and their nuanced ideas about consent and data practices; findings also detail conditions for student trust of LA. We close by discussing higher education's obligation to educate students about their LA practices and include students in privacy decisions as a means to reestablish contextual integrity and protect intellectual privacy.

Literature Review

Approaches to Student Privacy

An unambiguous student right to privacy in relation to LA would be a guiding concept for practice and policy, but such a right has been "elusive" to identify (Pardo & Siemens, 2014, p. 442). The lack of clarity in this conceptual space is not unexpected given privacy's multifaceted and complex characteristics. Pinning down privacy's definition and its intrinsic and instrumental value has been and continues to be an ongoing academic endeavor (Solove, 2005; Wu, Vitak, & Zimmer, 2019). Conceptions of privacy ebb and flow as new sociotechnical systems emerge, like LA, which present new privacy challenges for scholars and practitioners alike.

The literature recognizes that privacy must be considered as part of the success calculus of LA initiatives, but the treatment of privacy has at times been thin and underrepresented (Drachsler et al., 2015; Gašević, Dawson, & Jovanović, 2016). Some research promotes privacy principles to guide LA practice, such as those related to transparency, consent, and data ownership (see Drachsler & Greller, 2016; Pardo & Siemens, 2014); while others address ethical concerns that relate to privacy, like fairness and justice (see Scholes, 2016; Willis, Slade, & Prinsloo, 2016). However, neither of these literature groups attend to privacy's deep theoretical roots.

With the exception of a small group of scholars, the LA literature tends to assert that student privacy is valuable without explaining why. Heath's (2014) framework suggests that the theory of contextual integrity by Helen Nissenbaum (2010) can help define the normative value of privacy and identify when privacy expectations are violated in the context of LA. Hoel and Chen (2016) have also argued that contextual integrity holds significant promise for understanding and analyzing LA's privacy issues. Rubel and Jones (2016) state that privacy is "important as a function of autonomy" insofar that some want privacy and that privacy is necessary for autonomy. Further, they claim that claim that reductions of privacy (and autonomy) restrict human flourishing. Rubel and Jones argue that LA should not unjustifiably invade students' privacy as such actions have a knock-on effect of limiting their capability to pursue an educational program according to their own interests. There is another privacy framework that holds promise for LA, which is intellectual privacy, a theory that compliments the existing recommendations.

The Value of Intellectual Privacy to Higher Education

Before exploring intellectual privacy, it is useful to establish proposed benefits of higher education. There are a number of conceptualizations including, but not limited to, maximizing students' economic productivity; serving as an institution of justice, which is to say that universities provide opportunities for those who are typically disadvantaged by society; and serving society by creating new knowledge (Brighouse & McPherson, 2015). Intellectual privacy is in harmony with and supportive of an entirely different conceptualization of higher education: the value of a university education is that it provides students an opportunity to develop and test their creative and critical functions. Constructively participating in a pluralistic society requires one to have “complex skills and virtues such as those that enable deliberation and respect for reasonable differences” (Gutmann & Ben-Porath, 2015, p. 863). The ends to which developing such “skills and virtues” are directed include enabling students to “secure the basic liberties and opportunities of individuals, the collective capacity of individuals to pursue justice, and mutual respect in the face of disagreement” (Gutmann & Ben-Porath, 2015, p. 864). Growth in these areas is conditional. Gereluk (2018, p. 178) reminds us that “creativity, innovation, risk, and imagination...require space for deliberation and thought,” which higher education provides.

If these important skills and virtues are to develop, students need protected educational spaces and experiences that allow for “intellectual contemplation, idea generation and speech acts” (Jones & VanScoy, 2019, p. 1334) without unjustifiable intrusions or influence. Neil Richards’ (2012; 2015) theory of intellectual privacy explains that:

> New ideas often develop best away from the intense scrutiny of public exposure; that people should be able to make up their minds at times and places of their own choosing; and that a meaningful guarantee of privacy—protection from surveillance or interference—is necessary to promote this kind of intellectual freedom. (Richards, 2012, p. 1946)

Intellectual privacy provides the protections necessary when “we’re thinking reading, and speaking with confidants before our ideas are ready for public consumption” (Richards, 2015, p. 95). Contemplating existing ideas, considering belief systems, and testing the boundaries of both: these are all things that higher education institutions strive to do in the process of educating students. Intellectual privacy protects individuals from undue influence, enables them to think freely, provides them the necessary conditions for personal and intellectual growth, and ultimately serves as a means to their constructive participation in society.

While intellectual privacy can exist on its own as a descriptive approach to student privacy, we argue that it carries normative weight when combined with Nissenbaum’s (2010) framework of contextual integrity. Since intellectual privacy is compatible with higher education’s teleological foundations, it serves a role in determining the appropriate “transmission, communication, transfer, distribution, and dissemination” (Nissenbaum, 2010, p. 140) of information, or its flow characteristics. Intellectual privacy-sensitive informational norms have developed over time, which determine social expectations of flow, as well as legal and technological constraints. For instance, institutions have developed course- and institutional-level policies to protect their students’ privacy, as well as social expectations about what is right and wrong about particular student data and information practices. These policies prescribe specific actions, as is often the case with data stewardship committees, and they dictate consequences that will come from mishandling data (e.g., investigative procedures and repercussions, like dismissal). When new sociotechnical systems, such as LA, deny students’ ability to develop necessary skills and virtues, either by outright invading students’ privacy or limiting their intellectual freedom, they represent prima facie challenges to the contextual integrity of higher education. Not all changes in student information flows raise contextual integrity problems and, by extension, intellectual privacy concerns. For instance, legitimate data access and use practices, such as assessment, would be allowable assuming that such practices do not chill or unjustifiably influence a student's intellectual behaviors. However, LA and related educational data mining practices are not typical assessment activities. In the following sections of the literature review, we highlight how LA practices affect student data and information flows and bring about intellectual privacy concerns.

Institutional Flows of Student Data

Universities see value in the data they amass as students interact with their systems and disclose information about themselves. Institutional interest in exploiting data science methods, generally, and LA's potential, specifically, has led to changes in flows of student data and information in ways that are at times both novel and alarming. Since around 2010, mainstream LA has developed a set of standard practices that build from institutional data. Instructors use LA to visualize student progress in learning management systems using data dashboards (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). Advisors use predictive measures to identify a student's success probability in a course or program (Jones, 2019a). Institutional researchers build complex models to measure and predict student retention and graduation rates (Crisp, Doran, & Salis Reyes, 2017; Essa & Ayad, 2012; York, Gibson, & Rankin, 2015). And some academic librarians use LA to identify relationships between information access, use, and services with learning outcomes (Oakleaf & Kyrillidou, 2016; Oakleaf, Whyte, Lynema, & Brown, 2017). Beyond this small core of common practices has emerged some boundary-pushing initiatives, in so doing they explore new possibilities of what can be done with student data.

It is increasingly common for universities to use geolocation-enabled applications and devices, mobile or otherwise, to track student movements (Straumsheim, 2013). Sometimes this tracking is done for attendance-tracking purposes, while at other times institutions monitor student movements for seemingly benign reasons, like coaxing students to attend athletic events (Harwell, 2019; Witz, 2019). Similarly, institutions have been examining student movements to examine campus usage, social network development, and retention by exploring the data points created by student identification card swipes and the RFID chips embedded within (Blue, 2018; Meyer, 2018; Parry, 2012).

Universities have also started exploring what role, if any, voice assistants (e.g., Amazon's Alexa) can play in the educational experience (Ellis, 2018). While this technology may prove useful for students—and for the institutions who believe it can help with retention issues (Miles, 2019)—devices like the Amazon Echo are always on (Hobohm, 2019). This is especially concerning for students whose dorm rooms come equipped with the smart voice assistants, but would rather not risk the privacy intrusion (Price, 2019).

Facial recognition technology is becoming more prominent, as well. With school shootings in the United States increasing in number and frequency on elementary, secondary, and college campuses, school leaders have justified their adoption of facial recognition technology on security grounds (Heilweil, 2019). In higher education, the technological foundation for such tools is nearly perfect. Many institutions have closed-circuit televisions to support their campus security practices. Additionally, an institution typically requires its students to have their face photographed for identification purposes (e.g., student ID cards, online profiles), which are stored in and queriable from an institutional database. Facial recognition applications therefore have the static and dynamic data they need to identify a member of an institution's community—or a threat to its safety (Hernandez & Ehern, 2019). Andrejevic and Selwyn (2019) point out, however, that security is but one affordance of facial recognition. Knowing who students are enables institutions to authenticate learners via webcams in online learning environments. And when that authentication is paired with geolocation data, institutions can also automate attendance tracking. Finally, they argue that there is growing interest to study a student's "facial actions" to measure their classroom engagement and emotional state (Andrejevic & Selwyn, 2019, p. 5).

EdTech's Role in Student Data

One of the growing concerns associated with LA and related practices is how reliant institutions are on third parties to provide critical educational technology (edTech) services and infrastructures. To enable these technologies to be effective, institutions must provide access to and use of student data and information; some of which can be sensitive and would prove harmful if disclosed. One recent case demonstrates why the anxiety exists.

In 2019, the private equity firm Thoma Bravo began the process to acquire Instructure, the company behind the highly popular learning management system Canvas, for approximately $2 billion (Young, 2019). Instructure's CEO, Dan Goldsmith, had boisterously stated to investors earlier that year that the company had "the most comprehensive database on the educational experience in the globe.... no one else has those data assets at their fingertips to be able to develop those algorithms and predictive models," implicitly asserting that the financial value of Instructure was in part due to its access to student data (Young, 2020, para. 5). Given this statement and societal worry regarding the power technology companies wield over their users, vocal student advocates expressed their displeasure about the acquisition (Ethical EdTech, 2019).

The advocates' legitimate unease stemmed from their belief that Thoma Bravo perceived Instructure's value in the same way of its CEO and, as a result, would attempt to capitalize by selling its new data assets. In response, both Thoma Bravo and Instructure claimed that they would never share or sell student data (Young, 2020). While their statements may be true, there are good reasons to be troubled by edTech companies and the potential exploitation of student data. The companies collect considerable amounts of student data, which they use to fuel the development of remunerative products (Feldstein, 2016). And even if edTech companies do not sell their data assets, private equity firms will continue to look favorably at the data they hold. In fact, the Thoma Bravo acquisition was not the only notable acquisition in 2019. Advance Publications purchased Turnitin, a plagiarism detection company, for $1.75 billion (Johnson, 2019). The value of Turnitin existed in its massive trove of student intellectual property (e.g., written essays, projects, etc.), which by and large instructors provided to Turnitin for analysis, not students.

Weakening Privacy Protections

Institutions establish policies to protect their students' privacy, and the Family Educational Rights and Privacy Act (FERPA), the federal student privacy law in the United States, continue to safeguard students. But, both of these protective layers have weakened with the increasing use of LA. Christine Borgman (2018) argues that “in an ‘age of algorithms’ where datasets are in constant flux and can be disaggregated and reaggregated continuously for multiple analytical purposes, new approaches are sorely needed” in higher education (p. 384). Some institutions have responded to Borgman’s call to action by 1) recognizing the unique privacy issues associated with LA and educational data mining and 2) establishing protections matched to the emerging threats from predictive analytics and algorithms (see Tsai & Gašević, 2017; University of California, Berkeley, 2018; University of Hawaiʻi at Mānoa, 2018). While institutional policy can be responsive to sociotechnical advances, federal law has an earned reputation for reacting to such changes at a glacial pace; so it has been with federal student privacy law.

While receiving a number of amendments and clarifications since its introduction in 1974, FERPA is “ill-suited to anchor privacy protection in a big data world” (Polonetsky & Tene, 2015, p. 262). Elana Zeide (2017) asserts that “the statute no longer provides meaningful transparency to enable student...oversight of information accuracy or informed consent” (p. 503), which were some of the primary motivations for FERPA in the first place. The problem, she argues, is that information flows have changed so drastically with the evolution of old technology (e.g., databases) and the advent of new systems (e.g., facial recognition), that institutions no longer have the ability to track such flows, much less control them and inform students of how data about them are used, to what ends, and by whom. As such, a student’s “right of inspection, a right to amend inaccurate or misleading information, and a right to file a complaint where one’s rights are violated” is made impracticable given the complexities of today’s information flows (Rubel & Jones, 2016, p. 154).

Students and institutional actors may expect their institutional review boards (IRBs) to protect against foreseeable harms and act as an ethical check on LA. However, LA has created new information flow pathways that route around or make IRBs ineffectual. A major breakdown centers on whether LA is considered research, assessment, or part of quality improvement processes; the latter two can bypass ethical review entirely (Jones, 2019b). Even when ethical review does occur for research proper, IRBs may not fully anticipate the harms from allowing secondary uses of student data or the reidentification consequences from so-called “fishing expeditions” in large data sets (Willis et al., 2016). There have also been examples where edTech companies have used their access to student data to conduct research on unknowing student populations without consent, institutional approval, or IRB oversight (see Herold, 2018).

Student Perceptions and Expectations Regarding Learning Analytics

Information disclosure behaviors of traditional college-aged individuals have stirred claims that “young people don’t care about privacy” (Hargittai & Marwick, 2016, p. 3739). However, research on college-aged individuals’ privacy expectations and practices signals that students *do* care about their privacy, but their views, expectations, and practices are contextual and dependent on sociotechnical factors. Some researchers note an apparent “privacy paradox,” arguing that individuals often disclose sensitive information, yet claim to care about their privacy (Acquisti & Gross, 2006; Barnes, 2006; boyd & Hargittai, 2010). Other research attributes this supposed paradox to interfaces designed to mislead (e.g., dark patterns) (Waldman, 2020), learned helplessness (Hargittai & Marwick, 2016), or even a misleading sense of control (Brandimarte, Acquisti, & Loewenstein, 2012). Empirical evidence indicates that college-aged individuals are aware of privacy-promoting choices and express privacy preferences by strategically using a system's privacy affordances (see Blank, Bolsover, & Dubois, 2014; Shelton, Rainie, & Madden, 2015). Given the largely-held belief that students do not care about their privacy, it is necessary to investigate students' privacy perceptions in relation to specific technologies, such as LA. News articles have highlighted student dismay at the information their institutions keep on record, as well as the extent to which they track students' physical and digital movements (see Jones & McCoy, 2019; Rodrigo, 2020; Thorson, 2019; Vescera, 2019). Scholars are beginning to explore this area as well.

To better determine students' understanding, perceptions, expectations, and reactions to LA, researchers have primarily relied on focus groups and interviews (Bennett & Folley, 2019; Roberts, Howell, Seaman, & Gibson, 2016; Schumacher & Ifenthaler, 2018), mixed-methods (Ifenthaler & Schumacher, 2016; Prinsloo & Slade, 2015), and surveys (Slade, Prinsloo, & Khalil, 2019; Whitelock-Wainwright, Gašević, Tejeiro, Tsai, & Bennett, 2019). Only Whitelock-Wainwright et al. (2019) studied more than one institution, with all other studies restricted to a single institution. No studies have been conducted in the United States. Ifenthaler and Schumacher (2016), Prinsloo and Slade (2015), and Roberts et al. (2016) identified gaps in student awareness of data captured by higher education institutions. Bennett and Folley (2019) noted that students want to know the specific source of data being ingested and analyzed by LA dashboards and assurances that data were being used for educational purposes; these findings are similar to those of Slade et al., (2019). Additionally, Ifenthaler & Schumacher (2016) identified that students prefer that only academic data be shared for analysis; personal behaviors both online and off were out of bounds. Whitelock-Wainwright et al.'s (2019) findings highlighted that students expected their data to be handled ethically, which among other things included a requirement to be notified before data was used (which Slade et al.'s (2019) respondents wanted as well), that data access should be limited, and that data should be secured. Students remarked in multiple studies that they were concerned about a loss of autonomy and had concerns about unfair treatment arising from data dashboards and LA (Prinsloo & Slade, 2015; Roberts et al., 2016; Schumacher & Ifenthaler, 2018; Slade et al., 2019).

Emerging Research Questions

Further research is needed to fully explore student perceptions of the capture and usage of demographic data, physical and online behavior trails, and other non-academic data. Research is additionally needed on student perceptions within the United States. The study described hereafter addresses these gaps in the literature by conducting a multi-institution study within the United States using interviews guided by the following exploratory research questions: What privacy issues do undergraduate students perceive when discussing LA practices and initiatives? What are students' privacy expectations for LA data collected about them? How do students believed LA data should be used, shared, and protected? And finally, what are students' reactions to analytics in the context of complex educational environments?

Methods

The Constructivist Research Paradigm

This research is rooted in the naturalistic paradigm, specifically its branch of social constructivism (Charmaz, 2014; Lincoln & Guba, 1985), which argues that research subjects, such as students' perception and understanding of privacy, are socially constructed and influenced by respondents' experiences, relationships, values, and other aspects of their social milieu. Moreover, a social constructivist framework suggests that the development of students' privacy perceptions, as participants, are additionally impacted by the research process as participants when meanings and findings are co-constructed through interactions with the researcher "as the researcher aims to understand a phenomenon from the perspective of those experiencing it" (Costantino, 2008, p. 119). Methods such as semi-structured interviewing, as we employed in this study, are well suited to addressing phenomenological questions that seek interpretative, co-constructed answers.

The Research Context and Sample

The research team recruited undergraduate students over the age of 18 at eight United States higher education institutions:

1. Indiana University-Bloomington; Bloomington, Indiana
2. Indiana University-Indianapolis (IUPUI); Indianapolis, Indiana
3. Linn-Benton Community College; Albany, Oregon
4. Northwestern University; Evanston, Illinois
5. Oregon State University; Corvallis, Oregon
6. University of Illinois at Chicago; Chicago, Illinois
7. University of Wisconsin-Madison; Madison, Wisconsin
8. University of Wisconsin-Milwaukee; Milwaukee, Wisconsin

In order to maximize transferability of the research findings, efforts were made to include as many types of institutional profiles as possible based on Carnegie and National Center for Education Statistics classifications, as well as both public and private universities, institutions conferring four-year and two-year degrees, and institutions serving diverse student populations. Nevertheless, several limitations to this approach should be noted. Institutions were limited to institutions where research team members were affiliated or had standing research relationships, all the institutions are located in the Midwest United States except Oregon State University and Linn-Benton Community College, and several institutional categories are represented by only one institution such as private universities (Northwestern University), community colleges (Linn-Benton Community College), and Hispanic/Latino serving and Asian Pacific serving institutions (University of Illinois at Chicago).

The research team recruited students using a combination of methods, including random sampling, quota sampling, convenience sampling, and snowball sampling with a target of achieving 15 interviews at each institution and 120 in total. The interview target was chosen with this goal in mind: 15 interviews has been demonstrated to reveal over 90% of subsequently coded topics in qualitative interviews, or "saturation," the point where gathering additional data yields little new information (Guest, Bunce, & Johnson, 2006; Nielsen & Landauer, 1993). Participants were provided with a $10 Amazon gift card as an incentive to participate.

When available, the researchers utilized email lists provided by the institutions to contact a random sample of potential undergraduate participants, but participants were also recruited using flyers, listservs, and contacts with student groups. At the completion of each interview, researchers also encouraged participants to advertise the study to their peers. Because of this sampling approach, this study design therefore did not attempt to achieve representative results in the statistical sense. Instead, this study emphasizes dependability and confirmability by utilizing data analysis approaches that triangulate and corroborate findings across the eight institutions. Because conducting statistically significant tests comparing demographic groups was not a goal of this qualitative study, the research team chose not to systematically collect detailed demographic information.

Data Collection

Before beginning data collection, the semi-structured interview protocol was approved by each institution's IRB and classified as "exempt" from further review. The semi-structured interview protocol contained three core questions addressing students' overall perceptions and attitudes in relation to privacy and data collection practices, and, separately, five different sets of questions addressing the following themes:

1. Privacy
2. Data sharing and use
3. Data protections
4. Awareness of and reactions to LA
5. Libraries and LA

Each participant was asked the three common core questions and one discrete set of theme questions (see Appendix One) for a total of approximately ten questions. This approach was designed address the study's guiding research questions while also providing the breadth and flexibility required by an exploratory study. Each institution's researchers were asked to use each set of thematic questions three times in order to reach the overall target of 15 interviews for topic saturation within a theme. In addition to the core and theme protocol questions, interviewers asked probing and follow-up questions as necessary to elicit comprehensive responses and build researcher-participant rapport. The researchers gave participants flexibility regarding the format of the semi-structured interview: face-to-face, phone, or web conference. Researchers digitally recorded audio of all interviews for transcription purposes; a professional transcription company provided these services.

In total, the research team completed 112 semi-structured interviews. Recruitment issues prevented the team from meeting the target of 120 interviews. And after a thorough review of the transcripts, a further seven interview transcripts were removed from the final data set for quality control reasons due to variance from the protocol. The final data set therefore consists of 105 semi-structured interviews, which is sufficient to achieve adequate saturation of topics.

Data Analysis Procedures

The semi-structured interview transcripts were deposited into Dedoose, a collaborative, web-based qualitative data analysis application. Researchers wrote case summaries to accompany each transcript in order to provide “systematic, ordered” and succinct overviews of the facts of the interview related to key research questions (Kuckartz, 2014, p. 52) (see Appendix One). After completing the case summaries, the team collaborated to identify emerging themes and reflect on the data in order to inform the coding of the interview data.

Coding unfolded in three stages. For the first stage, researchers coded each transcript using the protocol question identification numbers. For the second stage, a sub-team of three researchers developed an exploratory codebook using a sample of the transcription data and an open coding technique. Researchers applied descriptive codes to data related to the research questions, which enabled the sub-team to understand the scope and breadth of the overall data set; the sub-team created 227 unique codes because of this generative process. The sub-team create a revised final codebook that included 79 codes grouped in seven thematic areas (see Appendix One). The full research team then completed closed coding using the codebook to ensure consistency across the coded data set.

We developed thematic findings for this article using quantitative and qualitative analysis techniques based on the closed coding results. Quantitatively, researchers analyzed nearly 11,000 code applications by examining code counts, code co-occurrence counts, and heatmaps to determine the most frequent combinations of codes and conspicuous absences in the code applications. From the quantitative analysis, the team collectively developed questions to drive further qualitative analysis strategies, including building combinations of protocol question codes and codebook codes to answer those questions by extracting and evaluating relevant transcript segments.

Evaluative Criteria

Social constructivist research is judged by unique evaluative criteria, including trustworthiness, dependability, and authenticity (Patton, 2008). The research team has attempted to meet these criteria by being transparent regarding its research design and related justifications as well as consciously and reflexively addressing their positionality and potential biases throughout this project. The team specifically developed interview questions designed to limit the potential for interviewer bias and to enable students to express their own views.

The findings presented below are clearly matched to the research questions, providing explanatory insights in the area studied. The study is dependable in part because the team developed an exhaustive data management strategy, documentation plan, and research infrastructure, which made the research process clear to the team and enabled constant communication about the project’s progress (see Jensen, 2008). Authenticity requires attending to “concerns about research that is worthwhile and thinking about its impact on members of the culture or community being researched” (James, 2008, p. 45). The research team has entered into this area of study directly because the findings could prove beneficial for the participants—the students whom they interviewed. Additionally, as we highlight in the findings, the research meets the authenticity criterion by helping students “develop greater understandings of the social context being studied” and by creating a “raised level of awareness” about LA and student privacy (James, 2008, p. 45).

Findings

Filling in Knowledge Gaps About Learning Analytics

Students lack data and privacy literacy or exhibit an inability to express privacy preferences due to opportunity and knowledge gaps regarding data practices in higher education. Participants did not fully consider their privacy as students because they signaled that the opportunity never presented itself. Consider some responses to the second core question, which asked students for their reactions to the fact that universities are increasing their data collection and analytic practices. One student stated, “I guess to be honest I’ve never really thought about it.” Several students explicitly stated that the call for participation in the study was the first time they had encountered the idea that their university was collecting and analyzing information about them.

Before interviewers provided examples of LA practices found in the literature, students often began a reflective process. Statements like “I think” or “now that I’m thinking about it” and “I assume” signaled the beginning of personal reflections on past knowledge of data practices outside of the higher education context and attempts to fill knowledge gaps regarding LA. The problem, as one student succinctly put it, was that “there’s less known about how universities use their information.” Even though students were uninformed about LA, they possessed the ability to parse important privacy-related questions about information flow characteristics.

Privacy theories and policies that regulate information flows emphasize information characteristics and transmission principles, such as data granularity, access, and use limitations. Students made similar analyses in their questioning. For instance, one participant asked about data sources and types, stating, “I don’t really know how the university is *getting my data* and *what kind of data* [emphasis added] they are gathering about me.” And another student raised questions about data usage and sharing, stating, “So that’s kind of new information to me…. I don’t know *how they’re using it* and *where they are taking it* [emphasis added]. So that part could definitely be concerning. Because I don’t know what they’re doing with it after they collect it.” Some of these reflections led students to consider at what point they had or someday might disclose personally identifiable data and information to their institution.

Participants often expressed an understanding that using campus technology systems and information networks could reveal their behaviors, often citing learning management systems (e.g., Canvas), WiFi networks, and identification swipe cards as technological artifacts and systems with which they often interacted. To a lesser but notable degree, some students identified their application for admission as a point of concern because they had “to surrender a significant amount of information,” for instance:

> My first encounter with this was the admissions process…. Like all the information colleges are getting about us. What are they doing about [it]? What happens when we get accepted, rejected? Who retains the information? What happens, basically? Because that’s like a lot of personal information you pour into the essay. It’s like your heart and soul.

This line of questioning often ended with a notable point of confusion: Why are institutions gathering and analyzing student data in the first place?

A Means to Unknown Ends

Students' lack of knowledge about LA led them to speculate about what it could achieve and whether those achievements justify data collection and use. Students assumed that LA would serve educational ends and that if this were so, this would be a "good" use, one that could "improve learning experiences," as a student explained. Similarly, another student stated "I feel that if they're using data altruistically in a sense to better the experience for every student as a whole, then I feel that I can see it as a positive endeavor… a win for everybody." Finally on this point, a student explained that "If they limit their purpose of collecting data solely for educational purposes, then I think it is fine…. I think it is very justifiable." What is particularly notable about these justificatory statements is their lack of granularity. Students did not—arguably because they *could not*—detail specific analytic practices that would actually improve learning experiences or serve the educational mission of an institution.

Students saw that they, their peers, and their institution could benefit from accessing and analyzing student information. Though the purposes of LA were unclear and the particulars unknown, students often used experiences with social media and eCommerce systems to envision how analytics could be used on their campus, for instance:

> Because if they had the intention of using my data to create better programs or better educational tools, then I'm all for it, you know. If there's data that can be tracked to say, "Oh, these students aren't learning well with this. We're seeing that this is actually really beneficial to the learning of our students." Then I'm like, "yeah, definitely. Please use this information, use this data to make the university better."

Even though students were fairly positive about LA prospects, this should not be misconstrued as an unlimited data access free-for-all.

Retaining Choice about Limiting Access

Students expressed nuanced arguments about when data access should be restricted, especially that their personal perspective on information access should not determine the privacy rights of their peers. While discussing the institution accessing their online searching behaviors, one student said:

> So, me personally, I don't search things that are really like too out there. I could see where someone else might do that and that would be like a problem; I definitely understand that. Maybe for me since I personally don't do anything that would embarrass me, I wouldn't be worried about the school seeing; it's not a problem for me… but for other students who might not be able to do what I'm doing, give them some privacy. Because who knows if [when] they go home that they even have WiFi, so they come to the school to utilize WiFi and do whatever they do in their free time.

Particularly telling about this excerpt are both the self-censorship in the student's use of institutional resources that could be surveilled, and the recognition that a lack of access to certain sociomaterial resources affect how peers use institutional resources. Students also recognized that in some cases their privilege afforded them more freedom to disclose data than others from less privileged and non-majority backgrounds. One student reflected on their comfort with information about a health diagnosis being known to the institution:

> I don't necessarily feel threatened by [my Autism status] being known and, like, that data being out there. But I understand that that could be very different for other students who don't pass as well as me, like students of color, low-income students, LGBT students, and other disabled students. There's plenty of reasons, like, they would feel the need to have that data protected. And I definitely understand that…. And I definitely have seen, like, the more marginalized identities a student holds, they tend to be, like, more suspicious of things like that and more wary of how their data is getting used.

Concerns about the inability to control access to mental-health care data stopped another student from seeking counseling at first:

> I could also see certain things that are tracked, maybe being a little embarrassing. I initially didn't go [to the counseling center] for a long time because I was embarrassed, because I knew that the university was going to be able to track that and look at my record and say, "Oh yeah, she's been going to counseling." And maybe if they wanted to, they could somehow find out what exactly it was that I was talking to the therapist about.

Students expected that uses of student information should map to the institution's responsibility to support its student body, and access should be limited to individuals who can claim to support students. It was unclear to students why other institutional actors would need to access data and information about them. Consistently, students reaffirmed that they should have a say over who has access to their data.

Neither Informed nor Consented

After identifying their knowledge gap about LA and institutional data practices, students probed why that gap exists. They concluded that their institution had failed to meaningfully inform them of its actions. As one student stated, "[higher education institutions] don't really give information to students [about] how they are using the student data." Without this information, as another student said, they have no idea "how deep [LA] goes" and "how far [institutions] will take it…. I just feel like we're being tracked at all times."

No student recalled explicitly consenting to campus data collection and use in any way (e.g., clickthrough or by signature). One student stated: "I don't know whether we ever signed anything when we became a student. Like, agreeing upon it. I mean, maybe we did and we just didn't realize it." Another student suggested that students "might have, like, signed something when they agreed to go to school." Students expressed varying degrees of suspicion and discomfort because of their uncertainty concerning informed consent; however, they suggested these negative feelings might subside with more transparent institutional practices. "I just feel like they could be a little more transparent about the fact that they are collecting it," said a student, "just making people aware of, like, what you're collecting and why." While most students discussed informed consent as an information practice that would serve their interests, one student notably suggested that it would be good for the institution as well:

> Something that would be helpful is having students sign off on saying, "Yeah. I recognize that my—what—my activity here at this university is being collected and that I'm okay with that." That would be probably, in my opinion, something that would be beneficial for both the university and students.

Should institutions not pursue any type of informed consent procedures, students conceded that they had little if any power to protest or change their institution's practices.

Differential Trust

Students compared institutional data practices and their privacy expectations with those around social media (e.g., Facebook) and eCommerce services (e.g., Amazon). This was both prompted and organic. Some interview questions specifically addressed these types of corporate entities, and students spontaneously incorporated them in answers to other questions as a point of reference. One significant point of comparison focused on trust: students clearly differentiated the low degree of trust they held toward corporate entities from the high degree of trust they felt toward their institution.

The single, most important factor informing students' differential trust was the for-profit or not-for-profit status of corporations and higher education institutions, respectively. As one student said, "I think you can trust [my institution]. I don't think I would be as wary as I am with, say, like bigger companies or something; they're not like for-profit [companies]." Another student stated that "I personally trust schools and universities more than these companies that are for-profit and I trust that they're going to use this information in a way that I feel more comfortable with, that doesn't try to take money out of my pocket." Other students added more nuance about institutional intent. One student said, "so, I'm more comfortable with them knowing things I have to say, or what I'm doing on the internet, I should say…. I trust that they're going to do *the right things* [emphasis added] with my data more so than companies." This student's expression of "the right things" signaled that students put trust in their institution because of a belief that colleges and universities are *moral* institutions. And because the perception was that universities would not capitalize on student data, they were more willing to share information about themselves for analytic ends.

Students ascribed a general ethical code to their institutions. They had "more faith" that their institutions would put students' needs and interests above their own, or at least consider students in their analytic practices. As one student said:

> When I think of a university, I definitely think of like a set of morals and principles that higher education holds. And I also, just from my experience at this university, I think that they really do have the students' best interests in mind…. When I think of the university, I think of, you know, we care about our students, we advocate for our students, we want students to learn. They're not all like that, but when I think of this one in general, I do hold them to a different kind of moral code.

This sense of trust and of institutional morality led many students to be comfortable with practices in a university setting that they were skeptical of in a corporate setting.

Data Sales as a Bright Line

Respondents were adamant that if higher education sold student data, their trust would dissolve because their privacy would be at risk. Since there are no publicly known instances of higher education institutions selling student data, this unambiguous firmness from respondents was unexpected. One participant said:

> I don't think the university should sell student data, because I don't think students want that and students did not put that data into university systems with the intent or the knowledge that it would ever be used for anything other than the purposes that [universities] were using it for, so I think that would be a betrayal of the university's students.

One possible explanation is that respondents are transferring their knowledge of data sales in eCommerce and social media environments to the educational environment, in so doing missing the significant legal and normative differences between these contexts.

Particularly interesting in this finding is that students view their relationship with their institution as transactional, just as they do with social media and eCommerce providers. As a student said, "I think it's kind of like the price I'm paying for it, I guess, like there's like information about myself that I'm giving up, but in exchange I'm getting more valuable resources." They recognize that they are paying for a service and that disclosing data is a part of establishing and maintaining that relationship. Another student said, "it's the school I go to and I know that they have to have that information, especially when I have to pay them." But unlike their relationships with social media and eCommerce providers, whom they know use and sell their data, they refuse to condone that practice from their institution. As one student put it, "when it comes to stuff like what [students] actually do, like let's say if the colleges go the same way as Facebook or Google or Amazon, then I completely disagree with that." Another student who believed data sales were already occurring said that such practices made them feel "insignificant" and "like a cog… not really a person anymore."

Discussion

An Obligation to Educate

LA advocates, privacy scholars, and data ethicists continue to debate a so-called "obligation to act" on student data using emerging analytic strategies (Kay, Korn, & Oppenheim, 2012; Prinsloo & Slade, 2017; Sclater, 2016; Willis, Campbell, & Pistilli, 2013). Those in favor of such an obligation argue that the it requires "the effective allocation of resources to ensure appropriate and effective interventions to increase effective teaching and learning" (Prinsloo & Slade, 2017, p. 46). "Knowing" that student data exists and *may* lead to downstream benefits creates an administrative and instructional "duty of care" arising from that knowledge (Kay et al., 2012, p. 5). Some even argue that this obligation supersedes a student's privacy expectations and preferences given that educational and even health benefits (e.g., preventative suicide detection) could occur and institutions could enjoy greater legal protections (Sclater, 2016). Nevertheless, similar "duty of care" arguments do not typically extend to include experimental treatments or interventions with unknown risks—even in medical contexts—and LA is presently in this experimental stage, with unclear risks and benefits.

A further moral problem with this formulation of an obligation to act is that it privileges the institutional viewpoint and all its byzantine political, economic, and social interests. It treats students as subjects, not autonomous individuals with their own values and interests (Rubel & Jones, 2016), which is especially problematic given the power imbalance between institutions and students. In so doing, it denies students both choice and voice. Our findings indicate that students hold positive viewpoints towards analytic practices and generally expect that institutions will collect and analyze student data. However, the findings also reveal that despite students' lack of specific knowledge about LA and poor recollection of LA-related consent practices or policies, they desire to have their privacy protected and preferences heard.

Institutions rarely acknowledge—much less privilege—student privacy preferences, in part to prioritize institutional interests. Until institutions meaningfully engage with students to repair the current information and power asymmetries between themselves and students (see Prinsloo & Slade, 2016), the purported obligation to act on LA leans toward the institution actors as self-interested, paternalistic, and unjustifiable. Jeffrey Alan Johnson (2017) came to a similar conclusion in a case study analysis of LA practices, arguing:

> [The analytics] model fails because it is structurally unjust; it oppresses and dominates students. Students' self-determination is undermined by organizational forms that establish paternalistic—literally, *in loco parentis*—authority over them. Using this authority, students' self-development is subordinated to the needs of institutions. (p. 21)

Colleges and universities, first and foremost, are obligated to educate their students about institutional data practices and how student data are and may be used, including potential harms as well as possible benefits. After educating students, they must also be willing to live with the plausible outcome that students do not want to be subjected to unjustifiable data collection and analysis.

Protecting Intellectual Privacy and Maintaining Contextual Integrity

Findings indicate that students are concerned about real and potential data mining and analytic practices to the point that they challenge their expectations of privacy. And in some cases, students reflect on and are willing to change their behaviors in light of knowledge of these practices. These findings demonstrate that LA presents very real challenges to intellectual privacy and contextual integrity. If, as Nissenbaum (2010, p. 140) writes, "[privacy expectations are] preserved when informational norms are respected and violated when informational norms are breached," then the study's findings clearly demonstrate that colleges and universities need to make concerted efforts to reestablish normative alignment in concert with student expectations. Like much of the extant literature on student privacy perspectives (see Bennett & Folley, 2019; Ifenthaler & Schumacher, 2016; Prinsloo & Slade, 2015; Roberts et al., 2016; Schumacher & Ifenthaler, 2018; Slade et al., 2019; Whitelock-Wainwright et al., 2019), the findings reinforce that students 1) are *not* noncommittal toward their privacy, 2) want to be a part of the decisionmaking process regarding their privacy, and 3) have particularized, multifaceted views regarding information access, control, and ownership (e.g., no sales of "my" data).

Even though higher education's LA practices lack integrity (with regard to privacy), all is not lost—just in need of careful consideration and strategic action to promote student privacy and informed consent. Solove (2008, p. 65) reminds us that "privacy is a condition we create, and as such, it is dynamic and changing." To begin rebuilding contextual integrity, it is crucial that institutions adopt a collaborative design process that genuinely and strategically includes the student voice. Co-designing policies, practices, and technologies will provide a constant feedback loop to universities and their representatives, and persistently force these representatives to justify their actions and be explicit about their goals. In so doing, it is likely that universities will be more attuned to students' privacy expectations and students will be more supportive and trusting of LA initiatives.

Conclusion

Higher education institutions have embarked on LA practices largely without consulting their students. To establish a baseline understanding about student knowledge, attitudes, and beliefs about privacy in relation to LA, we performed over 100 interviews with undergraduate students at eight higher education institutions across the United States. We discovered significant gaps in student knowledge. Respondents were largely unaware of LA practices, had very little idea how much data is collected about them, what is done with it, or how such data collection and practices could harm them. Respondent attitudes toward privacy could be quite nuanced. They saw potential benefits in some LA practices, but clearly expressed the desire to know about concomitant data collection and data use, and frowned on data sales and some data sharing.

Limitations of this work need to be noted. Our use of convenience sampling creates two issues. First, the eight institutions serving as sampling sites did reflect diverse types of institutions. But, with the exception of Linn-Benton Community College, the researchers sampled at their home institution, which weakened geographic diversity. Second, it is plausible that students with greater interest in privacy or analytics were more likely to participate. Another limitation concerns the lack of demographic data. Not seeking this data from our respondents limited our ability to draw thematic conclusions about demographic subgroups; the reported findings reflect only the larger population: undergraduates. However, this limitation was by design as the team planned in a second phase of the research to conduct a larger random-sample survey project shortly after this project's completion. The survey, which the team drafted before this article's publication, includes common demographic markers (e.g., self-identified gender, race, age, etc.) with academic characteristics (e.g., program of study, class standing, etc.).

Student privacy is just *one* of several complex LA issues that universities need to address, and it is not enough to simply seek student input on their privacy expectations. Although that input is needed and useful, it is limited by students' lack of data and privacy literacies that would enable them to parse larger systemic issues associated with data aggregation and analysis. For instance, while they see benefits in receiving personalized education and just-in-time services based on predictive scores, they were unable to express how such analytics could create harmful intellectual filter bubbles (Bozdag, 2013), lead to unfair panoptic sorts (Gandy, 1993), or deleteriously limit their educational experience in a critical time of personal, professional, and academic growth (Rubel & Jones, 2016). Ultimately, the responsibility rests with higher education institutions to carefully and fully consider the very real moral and ethical problems that LA creates. To aid their efforts, student-focused research needs to continue to dive into the particulars of granular problems such as data ownership, informed consent, trust, transparency, and others.

Acknowledgments

This project was made possible in part by the Institute of Museum and Library Services (LG-96-18-0044-18). The views, findings, conclusions or recommendations expressed in this paper do not necessarily represent those of the Institute of Museum and Library Services. The team thanks its research assistants for their support: Tanisha Afnan (Indiana University-Indianapolis), Aubree Tillett (University of Wisconsin-Milwaukee), Susan Rowe (Oregon State University), Ashwed Patil (Indiana University-Bloomington), and Donovan Pogue (Indiana University-Bloomington). The team also expresses its thanks to the anonymous peer reviewers for their insightful and constructive feedback. A version of this article was first published in the 2019 proceedings of the Association of College and Research Libraries' annual meeting.

References

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. *PET 2006: Privacy Enhancing Technologies*, 36–58. https://doi.org/10.1007/11957454_3

Andrejevic, M., & Selwyn, N. (2019). Facial recognition technology in schools: Critical questions and concerns. *Learning, Media and Technology*, 1–14. https://doi.org/10.1080/17439884.2020.1686014

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday*, *11*(9). https://doi.org/10.5210/fm.v11i9.1394

Bennett, L., & Folley, S. (2019). Four design principles for learner dashboards that support student agency and empowerment. *Journal of Applied Research in Higher Education, 12*(1), 15–26. https://doi.org/10.1108/JARHE-11-2018-0251

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief* (ED-04-CO-0040, Task 0010). U.S. Department of Education, Office of Educational Technology. Washington, DC.

Blank, G., Bolsover, G., & Dubois, E. (2014). *A new privacy paradox: Young people and privacy on social network sites*. Presented at the Annual Meeting of the American Sociological Association, San Francisco, CA. Retrieved from https://papers.ssrn.com/abstract=2479938

Blue, A. (2018, March 7). Researcher looks at 'digital traces' to help students. *University of Arizona News*. Retrieved from https://uanews.arizona.edu/story/researcher-looks-digital-traces-help-students

Borgman, C. L. (2018). Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, *33*(2), 365–412. https://doi.org/10.15779/Z38B56D489

boyd, d., & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, *15*(8). https://doi.org/10.5210/fm.v15i8.3086

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science*, *4*(3), 340–347. https://doi.org/10.1177/1948550612455931

Brighouse, H., & McPherson, M. (Eds.). (2015). Introduction: Problems of morality and justice in higher education. In *The aims of higher education: Problems of morality and justice* (pp. 1–6). Chicago, IL: University of Chicago Press.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Los Angeles, CA: SAGE Publications.

Cooper, A. (2012). *What is analytics? Definition and essential characteristics* (Volume 1, Number 5; CETIS Analytics Series). JISC. http://publications.cetis.org.uk/2012/521

Costantino, T. E. (2008). Constructivism. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 117–120). Thousand Oaks, CA: SAGE Publications, Inc.

Crisp, G., Doran, E., & Salis Reyes, N. A. (2017). Predicting graduation rates at 4-year broad access institutions using a Bayesian modeling approach. *Research in Higher Education*, *59*(2), 133–155. https://doi.org/10.1007/s11162-017-9459-x

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, 89–98. https://doi.org/10.1145/2883851.2883893

Drachsler, H., Hoel, T., Scheffel, M., Kismihók, G., Berg, A., Ferguson, R., Chen, W., Cooper, A., & Manderveld, J. (2015). Ethical and privacy issues in the application of learning analytics. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 390–391. https://doi.org/10.1145/2723576.2723642

Ellis, L. (2018, August 2). Hey, Alexa, should we bring virtual assistants to campus? *The Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/Hey-Alexa-Should-We-Bring/244129

Essa, A., & Ayad, H. (2012). Improving student success using predictive models and data visualisations. *Research in Learning Technology*, *20*(12), 58–70. https://doi.org/10.3402/rlt.v20i0.19191

Ethical EdTech. (2019). Letter to Instructure. Retrieved from https://ethicaledtech.info/wiki/Meta:Letter_to_Instructure

Feldstein, M. (2016, November 10). Popular discussion platform Piazza getting pushback for selling student data. Retrieved from E-Literate website: https://eliterate.us/popular-discussion-platform-piazza-getting-pushback-selling-student-data/

Gandy Jr., O. H. (1993). *The panoptic sort: A political economy of personal information*. Boulder, CO: Westview Press.

Gašević, D., Dawson, S., & Jovanović, J. (2016). Ethics and privacy as enablers of learning analytics. *Journal of Learning Analytics*, *3*(1), 1–4. https://doi.org/10.18608/jla.2016.31.1

Gereluk, D. (2018). Flourishing and wellbeing in the academy: A capabilities approach. *Philosophical Inquiry in Education*, *25*(2), 141–187. https://journals.sfu.ca/pie/index.php/pie/article/view/1063

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Gutmann, A., & Ben-Porath, S. (2014). Democratic education. In M. T. Gibbons, D. Coole, E. Ellis, & K. Ferguson (Eds.), *The Encyclopedia of Political Thought* (pp. 863–875). Chichester, UK: Wiley-Blackwell.

Hargittai, E., & Marwick, A. (2016). "What can I really do?" Explaining the privacy paradox with online apathy. *International Journal of Communication*, *10*(2016), 3737–3757. Retrieved from https://ijoc.org/index.php/ijoc/article/view/4655

Harwell, D. (2019, December 24). Colleges are turning students' phones into surveillance machines, tracking the locations of hundreds of thousands. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2019/12/24/colleges-are-turning-students-phones-into-surveillance-machines-tracking-locations-hundreds-thousands/

Heath, J. (2014). Contemporary privacy theory contributions to learning analytics. *Journal of Learning Analytics*, *1*(1), 140–149. https://doi.org/10.18608/jla.2014.11.8

Heilweil, R. (2019, December 20). Schools are using facial recognition to try to stop shootings. Here's why they should think twice. *Vox*. Retrieved from https://www.vox.com/recode/2019/12/20/21028124/schools-facial-recognition-mass-shootings

Hernandez, E., & Ehern, E. (2019, May 27). CU Colorado Springs students secretly photographed for government-backed facial-recognition research. *The Denver Post*. Retrieved from https://www.denverpost.com/2019/05/27/cu-colorado-springs-facial-recognition-research/

Herold, B. (2018, April 17). Pearson tested 'social-psychological' messages in learning software, with mixed results. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/DigitalEducation/2018/04/pearson_growth_mindset_software.html

Hobohm, T. (2019, April 29). Alexa, why are you here? *The Mercury*. Retrieved from https://utdmercury.com/alexa-where-are-you-here/

Hoel, T., & Chen, W. (2016). Privacy-driven design of learning analytics applications: Exploring the design space of solutions for data sharing and interoperability. *Journal of Learning Analytics*, *3*(1), 139–158. https://doi.org/10.18608/jla.2016.31.9

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, *64*(5), 923–938. https://doi.org/10.1007/s11423-016-9477-y

James, N. (2008). Authenticity. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 45–45). Thousand Oaks, CA: SAGE Publications.

Jensen, D. (2008). Dependability. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 209–209). Thousand Oaks, CA: SAGE Publications, Inc.

Johnson, J. A. (2017a). *Structural justice in student analytics, or, the silence of the bunnies*. Presented at the Digital Sociology Mini-Conference Eastern Sociological Society Annual Meeting, Philadelphia, PA. Retrieved from https://www.the-other-jeff.com/2017/02/structural-justice-in-student-analytics-or-the-silence-of-the-bunnies/

Johnson, S. (2019, March 6). Turnitin to be acquired by Advance Publications for $1.75B. *EdSurge*. Retrieved from https://www.edsurge.com/news/2019-03-06-turnitin-to-be-acquired-by-advance-publications-for-1-75b

Jones, K. M. L. (2019a). Advising the whole student: EAdvising analytics and the contextual suppression of advisor values. *Education and Information Technologies*, *24*(1), 437–458. https://doi.org/10.1007/s10639-018-9781-8

Jones, K. M. L. (2019b). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of Educational Technology in Higher Education*, *16*, 1–22. https://doi.org/10.1186/s41239-019-0155-0

Jones, K. M. L., & McCoy, C. (2019). Reconsidering data in learning analytics: Opportunities for critical research using a documentation studies framework. *Learning, Media and Technology, 44*(1), 52–63. https://doi.org/10.1080/17439884.2018.1556216

Jones, K. M. L., & VanScoy, A. (2019). The syllabus as a student privacy document in an age of learning analytics. *Journal of Documentation, 75*(6), 1333–1355. https://doi.org/10.1108/JD-12-2018-0202

Kay, D., Korn, N., & Oppenheim, C. (2012). Legal, risk and ethical aspects of analytics in higher education. *CETIS Analytics Series*. Retrieved from https://publications.cetis.org.uk/2012/500

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software*. SAGE.

Lane, J. E., & Finsel, B. A. (2014). Fostering smarter colleges and universities: Data, big data and analytics. In J. E. Lane (Ed.), *Building a Smarter University: Big Data, Innovation, and Analytics. Critical Issues in Higher Education* (pp. 3–26). Albany, NY: SUNY Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: SAGE Publications.

Meyer, D. (2018, March 13). An American university is spying on students to predict dropouts. *Fortune*. Retrieved from http://fortune.com/2018/03/13/university-arizona-catcard-big-data-dropouts/

Miles, K. (2019, December 27). Should colleges really be putting smart speakers in dorms? *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/614937/colleges-smart-speakers-in-dorms-privacy/

Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big Data, big analytics: Emerging business intelligence and analytic trends for today's businesses*. Hoboken, NJ: Wiley.

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, 206–213. ACM. https://doi.org/10.1145/169059.169166

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford Law Books.

Nunn, S., Avella, J. T., Kanai, T., & Kebritchi, M. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning, 20*(2), 13–29. https://doi.org/10.24059/olj.v20i2.790

Oakleaf, M., & Kyrillidou, M. (2016). Revisiting the academic library value research agenda: An opportunity to shape the future. *The Journal of Academic Librarianship*, *42*(6), 757–764. https://doi.org/10.1016/j.acalib.2016.10.005

Oakleaf, M., Whyte, A., Lynema, E., & Brown, M. (2017). Academic libraries & institutional learning analytics: One path to integration. *The Journal of Academic Librarianship*, *43*(5), 454–461. https://doi.org/10.1016/j.acalib.2017.08.008

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Parry, M. (2012, July 18). Colleges awakening to the opportunities of data mining. *The New York Times*. Retrieved from https://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html

Patton, M. Q. (2008). Evaluation criteria. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 302–303). Thousand Oaks, CA: SAGE Publications, Inc.

Polonetsky, J., & Tene, O. (2015). Who is reading whom now: Privacy in education from books to MOOCs. *Vanderbilt Journal of Entertainment and Technology Law, 17*(4), 927–990.

Price, M. (2019, August 13). Alexa, time for class: How one university put an Echo Dot in every dorm room. *CNET*. Retrieved from https://www.cnet.com/features/alexa-time-for-class-how-one-university-put-an-echo-dot-in-every-dorm-room/

Prinsloo, P., & Slade, S. (2015). Student privacy self-management. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 83–92. https://doi.org/10.1145/2723576.2723585

Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics, 3*(1), 159–182. https://doi.org/10.18608/jla.2016.31.10

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: The obligation to act. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 46–55. https://doi.org/10.1145/3027385.3027406

Richards, N. (2015). *Intellectual Privacy: Rethinking Civil Liberties in the Digital Age*. Oxford, UK: Oxford University Press.

Richards, N. M. (2012). The dangers of surveillance. *Harvard Law Review*, *126*, 1934–1965. Retrieved from https://harvardlawreview.org/2013/05/the-dangers-of-surveillance/

Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student attitudes toward learning analytics in higher education: "The Fitbit version of the learning world." *Frontiers in Psychology*, *7*, 1–11. https://doi.org/10.3389/fpsyg.2016.01959

Rodrigo, C. M. (2020, January 14). Digital rights group, students team up against facial recognition tech on college campuses. *The Hill*. Retrieved from https://thehill.com/policy/technology/478108-digital-rights-group-students-team-up-against-facial-recognition-tech-on

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, *32*(2), 143–159. https://doi.org/10.1080/01972243.2016.1130502

Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development*, *64*(5), 939–955. https://doi.org/10.1007/s11423-016-9458-1

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, *78*, 397–407. https://doi.org/10.1016/j.chb.2017.06.030

Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics*, *3*(1), 16–42. https://doi.org/10.18608/jla.2016.31.3

Shelton, M., Rainie, L., & Madden, M. (2015). *Americans' privacy strategies post-Snowden*. Retrieved from Pew Research Center website: https://www.pewinternet.org/2015/03/16/americans-privacy-strategies-post-snowden/

Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 4–8. https://doi.org/10.1145/2330601.2330605

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, 235–244. https://doi.org/10.1145/3303772.3303796

Solove, D. J. (2005). A taxonomy of privacy. *University of Pennsylvania Law Review*, *154*(3), 477.

Solove, D. J. (2008). *Understanding privacy*. Cambridge, MA: Harvard University Press.

Straumsheim, C. (2013, October 18). Before the fact. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2013/10/18/u-kentucky-hopes-boost-student-retention-prescriptive-analytics

Thorson, A. (2019, November 15). VCU plans to track student attendance with new pilot program. *WFXR*. Retrieved from https://www.wfxrtv.com/news/vcu-plans-to-track-student-attendance-with-new-pilot-program/

Tsai, Y.-S., & Gašević, D. (2017). Learning analytics in higher education --- challenges and policies: A review of eight learning analytics policies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 233–242. https://doi.org/10.1145/3027385.3027400

University of California, Berkeley. (2018). *Learning data principles*. https://rtl.berkeley.edu/learning-data-principles

University of Hawaiʻi at Mānoa. (2018). *Resolution supporting learning data privacy principles and practices*. https://docs.google.com/document/d/1bqJTZ4tCK3SFdsS4rXVNK5xlS87g0ajfE3EYk3u0Zh8

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363

Vescera, Z. (2019, March 27). Canvas is tracking your data. What is UBC doing with it? *The Ubyssey*. Retrieved from https://www.ubyssey.ca/features/double-edged-sword/

Waldman, A. E. (2020). Cognitive biases, dark patterns, and the 'privacy paradox.' *Current Opinion in Psychology*, *31*, 105–109. https://doi.org/10.1016/j.copsyc.2019.08.025

Whitelock-Wainwright, A., Gašević, D., Tejeiro, R., Tsai, Y.-S., & Bennett, K. (2019). The student expectations of learning analytics questionnaire (SELAQ). *Journal of Computer Assisted Learning, 35*(5), 633–666. https://doi.org/10.1111/jcal.12366

Willis III, J. E., Campbell, J. P., & Pistilli, M. D. (2013). Ethics, big data, and analytics: A model for application. *EDUCAUSE Review Online*. https://er.educause.edu/articles/2013/5/ethics-big-data-and-analytics-a-model-for-application

Willis III, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: A typology derived from a cross-continental, cross-institutional perspective. *Educational Technology Research and Development*, *64*(5), 881–901. https://doi.org/10.1007/s11423-016-9463-4

Witz, B. (2019, September 12). Orwellabama? Crimson Tide track locations to keep students at games. *The New York Times*. Retrieved from https://www.nytimes.com/2019/09/12/sports/alabama-tracking-app.html

Wong, B. T. M. (2017). Learning analytics in higher education: An analysis of case studies. *Asian Association of Open Universities Journal, 12*(1), 21–40. https://doi.org/10.1108/AAOUJ-01-2017-0009

Wu, P. F., Vitak, J., & Zimmer, M. T. (2019). A contextual approach to information privacy research. *Journal of the Association for Information Science and Technology, 71*(4), 485–490. https://doi.org/10.1002/asi.24232

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research & Evaluation*, *20*(5), 1–20. https://doi.org/10.7275/hz5x-tx03

Young, J. R. (2019, December 4). New ownership for an LMS giant: Private equity firm to buy Instructure for $2 billion. *EdSurge*. Retrieved from https://www.edsurge.com/news/2019-12-04-new-ownership-for-an-lms-giant-private-equity-firm-to-buy-instructure-for-2-billion

Young, J. R. (2020, January 17). As Instructure changes ownership, academics worry whether student data will be protected. *EdSurge*. Retrieved from https://www.edsurge.com/news/2020-01-17-as-instructure-changes-ownership-academics-worry-whether-student-data-will-be-protected

Zeide, E. (2017). The limits of education purpose limitations. *University of Miami Law Review, 71*, 493–526.

Appendices

Appendix One: Research Tools

The research team has compiled all relevant research tools associated with the project and made them publicly available at an Open Science Framework (OSF) repository. This was done in part to increase transparency around the design of the research and to support extensions from this work by others. Readers can access these materials using the following citation:

> Jones, K. M. L., Asher, A., Briney, K. A., Goben, A., Perry, M. R., Robertshaw, M. B., & Salo, D. (2019). Data Doubles: Research tools for Phase One interviews. https://doi.org/10.17605/OSF.IO/VBT93