Ethical and Administrative Policy Concerns about use of Big Data in K-12 Education

Priscilla M. Regan
Schar School of Policy and Government
George Mason University
November 2017
pregan@gmu.edu

Much of the ethical discussion about big data in education has been framed in terms of "privacy." This is not particularly surprising both because privacy is viewed as a multi-faceted concept with several different components and also because discussions about ethics and information technology in other sectors and over time have often been categorized under the value of privacy. I identify six separate concerns traditionally associated with privacy that are challenged by big data generally and also in the context of education.

I caution against a narrow approach that discusses these ethical issues in the context of the current legal framework and standard fair information principles. Instead a broader understanding of these ethical issues is necessary, as well as a realistic view of the management and administrative constraints in the K-12 environment.

Six Ethical Policy Concerns[1]

The *first* is that collection of information about an individual should take place with the knowledge of the individual and that the amount of information should be minimized to that which is required for the particular purpose for which it was collected. This is the classic information privacy concern that from a policy perspective has been addressed by the Fair Information Practice Principles (FIPPs) often summarized by notice, consent, choice and transparency. These principles are the basis of much privacy and data protection legislation around the world including in the United States in the Family Educational Rights and Privacy Act (FERPA) and the Children's Online Privacy Protection Act (COPPA).

Although many have questioned the effectiveness of the FIPPs approach more generally, there is almost universal agreement among privacy scholars and experts that the FIPPs approach is not appropriate in the big data environment. With big data there is more

[1] For a more complete discussion of these issues, see (available on request from author): Priscilla M. Regan, Jolene Jesse and Elsa Talat Khwaja, "Big Data in Education: Developing Policy for Ethical Implementation in the US and Canada" at the annual meetings of the American Society for Public Administration in Seattle, March 18-22, 2016. This research was supported by the Social Sciences and Humanities Research Council of Canada.

collection of information, by more parties, about more aspects of an individual's life, and with more granularity about that life. But the issue is not merely "more" or even the qualitative changes that quantity does not convey. The issue is also how much of big data collection takes place without the individual's awareness. As the President's Council of Advisors on Science and Technology (PCAST) noted in 2014 individuals "constantly *emit* into the environment information whose use or misuse may be a source of privacy concerns."[2]

With respect to education and big data, this issue of notice, consent and transparency becomes even more complicated than it does in other contexts both because records of children and hence the concerns of parents come into play and also because the educational relationship is mandatory, not voluntary. Educational technology firms usually do not generally have a direct contractual relationship with the students and parents but with the schools, school boards or teachers. Thus a policy that provides information and controls about the uses of big data are at least one step removed from the data subject and therefore more difficult for the data subject to "control."

Those advocating for the benefits of edtech in the K-12 environment acknowledge that there are privacy concerns but suggest that these can be addressed through regulations that provide notice and transparency. These groups include the Data Quality Campaign (DQC), the Future of Privacy Forum (FPF), the Consortium for School Networking (CoSN), the Student Privacy Resource Center (FERPASherpa), and the Software and Information Industry Association (SIIA). Among the activities of these organizations are the creation of "pledges" and "certifications" that educational technology companies and education leaders could sign on to by promising to adopt prescribed privacy practices. The Student Privacy Pledge, for example, was developed by FPF and SIAA as a way for educational technology companies to pledge to more open communication about their products and privacy safeguards and to encourage the adoption of practices that "meet or go beyond" federal regulations. The website claims 243 current signatories.[3] CoSN is also developing a "Trusted Learning Environment Seal" targeting "school system leaders" who have undergone the organization's certification programs to become "certified education technology leaders."[4] Finally, DQC also targets school leaders with information about communicating about the benefits of using data on student achievement, and on applicable privacy laws and protections through online training modules and awards for state and local officials who "have embraced a culture of data in service of students."[5]

---

[2] President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective* (May 2014), p. 38. Available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf, x

[3] More information about the Student Privacy Pledge may be found at https://studentprivacypledge.org/ accessed March 8, 2016.

[4] More information about the Trusted Learning Environment Seal may be found at http://www.cosn.org/about/news/national-education-organizations-launch-effort-build-%E2%80%98trusted-learning-environment%E2%80%99-us-1 accessed March 8, 2016.

[5] More information about the Data Quality Campaign and their Flashlight Awards may be found at http://dataqualitycampaign.org/success-stories/data-flashlight-awards/ accessed March 8, 2016.

Some question the efficacy of privacy pledges and certification. Natasha Sanger, reporting in the *New York Times Bit Blog* in February 2015, noted that a Student Privacy Pledge signature does not guarantee that companies have adopted the best encryption practices to protect student data on unsecured networks. Additionally, the education technology companies that sign the pledge, while promising to protect student data, do not commit to protecting teacher and/or parent data collected.[6] Others have raised issues of data privacy equity as well. While well-funded school districts might be able to afford well-designed education software and apps with top-of-the-line privacy and security protections, poorer school districts may find they rely more on free software from non-profits or fledgling startups that might not be able to afford the best data encryption measures, regardless of whether they have signed a pledge to do so.[7]

A *second* concern long associated with privacy is that individuals should be able to remain anonymous or obscure if they so choose to do so. But with an ever-increasing number of social relationships and practices becoming data points, it becomes more difficult for individuals to remain unidentified or unfindable. Algorithmic searches of datasets now can rather quickly eradicate what had been high transaction costs on finding meaningful information.[8] Most privacy and data protection laws cover "personal information" or "personally identifiable information" meaning that the information was directly associated with a particular individual. With big data, such distinctions are obviated as more and more bits of unidentified information can in effect be attached to a particular individual with just a bit of searching and analysis. With big data, anonymization of information about individuals becomes more difficult, if not impossible, as big data makes reidentifying data rather easy.[9]

Educational data are often stored in large, longitudinal data sets from which personally identifiable variables have been removed. These data sets are used for reporting purposes from the school to district to state or province and finally to the federal government. They are also used for research purposes to identify trends over time and to analyze factors that affect student performance. They have traditionally been referred to as aggregate,

---

[6] Sanger, Natasha, February 11, 2015, "Data Security Gaps in an Industry Student Privacy Pledge," New York Times Bit Blog available at http://bits.blog.ntimes.come/2015/02/11/data-security-gaps-in-an-industry-student-privacy-pledge/ accessed February 16, 2016.

[7] Sanger outlines instances of poor data encryption, and issues of equity are brought up in "From Mining to Minding Student Data," EdSurge, accessed March 8, 2016 at https://www.edsurge.com/research/special-reports/state-of-edtech-2016/k12_edtech_trends/data_privacy

[8] Woodrow Hartzog and Evan Selinger, "Big Data in Small Hands," *Stanford Law Review Online* (Sept. 3, 2013) 66:81-88 and Woodrow Hartzog and Evan Selinger, "Obscurity: A Better Way to Think about Your Data than Privacy," *Atlantic* (Jan. 17, 2013). Available at: http://www.theatlantic.com/technology/archive/2013/01/obscurity-a-better-way-to-think-about-your-data-than-privacy/267283/

[9] Latanya Sweeney, *Uniqueness of Simple Demographics in the US Population* (Laboratory for International Data Privacy, Working Paper LIDAP-WP4, 2000). Available at: http://dataprivacylab.org/projects/identifiability/index.html. Latanya Sweeney, Akua Abu, and Julia Winn, "Identifying Paticipants in the Personal Genome Project by Name," *Harvard University Data privacy Lab,* White Paper 1021-1 (April 24, 2013). Available at: http://dataprivacylab.org/projects/pgp/1021-1.pdf

anonymized data – but this tradition is being challenged in the era of big data. After reviewing the computer science and legal literatures on anonymity and reidentification, Paul Ohm concludes that: "Data can be either useful or perfectly anonymous but never both."[10] As a biomedical researcher notes: "I can't anonymize your genome without wiping out the information that I need to analyze."[11] Much the same holds true in the educational context.

A *third* concern involves the surveillance or tracking that provides more and more detailed information for big data analytics. In the area of big data and education, online testing and teaching programs monitor how long it takes students to answer a question or read a page – and often also capture key strokes or patterns of reading or responding that might indicate the thought processes of the student. The programs may also track where (home, school, computer lab) the student is working and what time of day – and often also record what other students are working on the same programs at that time. The results of all this tracking can be cross-matched with more traditional information about the student as well as new information from various devices (such as how much a student moves throughout the day or how much time a student spends on social networking sites) – and all of this may be fed into predictive analytics programs to determine student learning patterns, strengths and weaknesses, and advice about how best to personalize the learning environment for that student – and thus raise a *fourth* ethical concern regarding autonomy.

Big data, especially the analytics powered by big data, challenge individual autonomy, the individual's ability to govern his or her life as that individual thinks best. Big data algorithms jeopardize autonomy by leading people in certain directions. Ian Kerr and Jessica Earle distinguish among three types of predictions that affect autonomy: consequential predictions that allow individuals to act more in their self-interest and avoid unfavorable outcomes; preferential predictions that lead one to act in a way expected from the data; and preemptive predictions that are not based on the preferences of the actor but reduce the range of options available to the actor.[12] Tene and Polonetsky point to the dangers of predictive analysis including the perpetuation of old prejudices and the accentuation of social stratification.[13]

---

[10] Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," 57 *UCLA Law Review* 1701-1777, 1704 (2010).

[11] John Quackenbush quoted in Jonathan Shaw, "Why 'Big Data' is a Big Deal," *Harvard Magazine* (March/April 2014), 30-35, 74-75, p. 34. Available at: http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal

[12] Ian Kerr and Jessica Earle, "Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy, *Stanford Law Review Online* (Sept. 3, 2013) 66: 65-72.

[13] Omar Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," *Northwestern Journal of Technology and Intellectual Property* (2013) 11(5): 239-273 ,253. Available at: http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip

In the education environment, predictive analytics may lead to "tracking" of students thus foreclosing options that students may have selected for themselves or may have been well-suited to pursue but at a later point in their educational experience.[14]

Autonomy is thus related to a *fifth* privacy concern associated with big data, which involves traditional due process for individuals, the principle that individuals are treated fairly and equally and not discriminated against based on race, gender, age or other personal attributes – or based on factors of which they are not aware. Big data's use of mathematical algorithms and artificial intelligence to make predictions about individuals based on conglomerates of their information and that of others raises questions about treating individuals as individuals fairly, accurately, and in ways they can understand.[15] This concern involves issues of profiling and discrimination.

In the education environment, with its recognition of the importance of education to equal opportunity, there is a longstanding concern for not discriminating and for watching closely for subtle, as well as obvious, signs of discrimination. But with big data such subtle signs may be difficult to discern. For example, Ohm points out that "big data helps companies find a reasonable proxy for race."[16] But perhaps more troubling in education is that big data facilitates the creation of more refined, intersectional categories that discriminate among students in more insidious and harder to read ways. At a Data and Civil Rights Conference in 2014, these issues were explicitly addressed in one paper in which the authors pointed out: "the complexity of algorithmic analysis makes identification of bias and discrimination difficult;" the difficulty of reversing or avoiding "flawed algorithmic assessments;" the danger of self-fulfilling prophecies or prejudging students; and the risk of increasing stratification.[17]

A *sixth* issue that has long been part of the debate about privacy, especially information privacy, is the question of the ownership of data about an individual. Does the individual "own" the information or does the third party holding the information in a database? Although many privacy scholars question whether the property model provides a workable framework for talking about privacy,[18] the property rhetoric and rationales have become

---

[14] For a discussion of this issue, see (available on request from author): Priscilla M. Regan, Jolene Jesse and Elsa Talat Khwaja, "Big Data in the Education Arena: 21st Century Student Sorting and Tracking," at the 7th Biannual Surveillance and Society Conference in Barcelona Spain, April 20 to April 23, 2016.

[15] Danielle Keats Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washington Law Review* (2014) 89: 101-133.

[16] Paul Ohm, General Principles for Data Use and Analysis," in Julia Lane Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum (eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement.* New York: Cambridge University Press, 2014, pp. 96-111.

[17] Andrea Alarcon, Elana Zeide, Alex Rosenblat, Kate Wikelius, danah boyd, Seeta Pena Gangadharan, and Corrine Yu, "Data & Civil Rights: Education Primer," produced for Data & Civil Rights Conerence (October 30, 2014) available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2542268

[18] Paul M. Schwartz, "Property, Privacy, and Personal Data," *Harvard Law Review* 117(7):2055-2128 (May 2004) and Julie E. Cohen, "Examined Lives: Informational Privacy and the Subject as Object," *Stanford Law Review* 52: 1373-1438 (2000).

part of the policy discussion about big data, as they had been in earlier iterations of debates about privacy policy. In the education arena, student records are generally "owned" by the school or school district. The involvement of ed tech companies has somewhat muddied the question of ownership – depending on how contracts with these firms are written.

Management and Administrative Realities in K-12 Environment

With the advent of edtech applications in the K-12 environment, students are often working individually on a computer engaging with an educational challenge that is different than that of the student at the next computer. From the student's view, he or she is interacting with some software; the student is likely unaware of the complicated administrative and technological infrastructure that is behind that computer application. And in many cases, the teacher, parent, principal and school board – as well as the state education department – may also be unaware of this complex infrastructure. I believe this is a second important reason why current discussions of edtech and student privacy should not be framed solely in terms of the existing legal requirements but need to be broadened to take into account the various technical, legal and administrative constraints and loopholes that exist as schools and others contract with edtech companies.[19]

Two concerns in particular stand out in this complex environment.

First, edtech companies appear to be marketing primarily to schools and teachers, emphasizing the educational opportunities of their products, and downplaying their privacy implications. As a result new policies need to address the totality of the relationship that edtech companies enter with schools and teachers.

A preliminary review of company websites[20] in 2015 revealed not only that companies are marketing primarily to schools and teachers but also that privacy is rarely highlighted in marketing and promotional materials, which predictably tend to highlight the benefits of technology and data-driven education. Uncovering privacy statements can sometimes take many mouse clicks with a confusing array of privacy statements for use of the website versus use of the software. While some companies will include information about their signature on certifications such as the Student Privacy Pledge, the US EU Safe Harbor Framework, TRUSTe Privacy Seals, FERPA compliance and the like, this is no guarantee of privacy compliance. Most privacy policies, once found, are written in somewhat inaccessible language and are relatively short. A 2014 *Politico* investigation found similar patterns in ed tech companies' policies and practices, taking particular note of their "legal jargon and fuzzy terminology," that companies "typically reserve the right to change the policy at any time," and that the information "may be

[19] For a more complete discussion of these, please see on request from the author: Priscilla M. Regan and Elsa Talat Khwaja, "Ethical Implementation of Big Data in Education: Policy and Practices in the US and Canada," Presented at the Law and Society Association Annual Conference, June 2017, Mexico City.

[20] We investigated the websites of the following ed tech vendors: Schoology.com; Edmentum; Remind; Edsby; PowerSchool SIS; Clever; Public Consulting Group Canada; SAS Enterprise Analytics for Education; McGraw-Hill Connect; LoudCloud Systems; Amplify; Tenmarks-Amazon; and Google for Education. Six of the 13 vendors signed the Student Privacy Pledge discussed in the next section.

subject to an entirely new privacy policy, if the company is sold – a common fate for a start-up."[21]

It appears that companies have not yet used data privacy and security as a marketable component of their software, or made it easy for schools, teachers, parents or students to make informed decisions about data use and ownership. The US Department of Education's Privacy Technical Assistance Center as addressed these concerns in a brief addressing requirements and best practices in protecting privacy while using online educational practices. Its response to the questions of whether or how FERPA and PPRA covered online educational services and protected student records in this environment was basically – "it depends. Because of the diversity and variety of online educational services, there is no universal answer…"[22]

In January 2015 the Department of Education released a "model terms of service" document outlining best practices in contracts and agreements with ed tech vendors and wording in such agreements that should be avoided:

> PTAC offers this guidance to schools and districts to help them evaluate potential TOS agreements, and to offer direction regarding terminology frequently used in these agreements. By understanding commonly used provisions, schools and districts will be better able to decide whether to consent to a Click--Wrap or other TOS agreement for online educational services and mobile applications. The best practice recommendations below may also assist providers by suggesting approaches that better protect student privacy.[23]

The document offers guidance on several ethical issues identified above. With respect to the possibility of re-identification of data, the model terms of service caution that TOS agreements should prohibit re-identification by the vendor and in any future data transfers and that the agreements should specify that de-identification "requires more than just removing any obvious individual identifiers, as other demographic or contextual information can often be used to re--identify specific individuals. Retaining location and school information can also greatly increase the risk of re-identification."[24] In addition, TOS agreements should:

- prohibit use of data or metadata to create profiles of students or parents for marketing purposes;
- specify that at the end of a contract data must be destroyed or returned to the school or school district;
- indicate that ownership of data remains with the school or school district.

---

[21] Stephanie Simon, "The big biz of spying on little kids," *Politico* (May 15, 2014), available at: http://www.politico.com/story/2014/05/data-mining-your-children-106676

[22] Department of Education, Privacy Technical Assistance Center, "Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices," PTAC-FAQ-3 (February 2014), available at: https://tech.ed.gov/wp-content/uploads/2014/09/Student-Privacy-and-Online-Educational-Services-February-2014.pdf

[23] Department of Education, Privacy Technical Assistance Center, "Protecting Student Privacy While Using Online Educational Services: Model Terms of Service," PTAC-FAQ-4 (January 2015) http://ptac.ed.gov/sites/default/files/TOS_Guidance_Jan%202015_0.pdf

[24] Ibid.

The document is written with the goal to ensure that TOS agreements are in compliance with FERPA and PPRA.

During 2015, there was some bipartisan congressional interest in student data privacy. Representatives Todd Rokita (R-IN) and Marcia Fudge (D-OH) introduced an amendment to the Family Educational Rights and Privacy Act (FERPA) with the goal to increase the federal government's enforcement authority over service providers that misuse student data (DQC, 2015, p. 2). The Senate also adopted an amendment, introduced by Orrin Hatch (R-UT) and Edward Markey (D-MA), to the existing Elementary and Secondary Education Act, which aimed to create a Student Data Privacy Policy Committee with responsibility for studying and providing recommendations on privacy safeguards and parental rights. 2015 also witnessed bills introduced independently of existing federal statute including the Student Digital Privacy and Parental Rights Act and the SAFE KIDS Act — both modeled somewhat after California's SOPIPA (discussed below) and designed to provide some regulation over online education service providers.[25]

Second, federal efforts should complement state efforts. States are beginning to address the appropriate roles of state boards of education, school districts, and school boards:

- In 2014, 32 bills charged state boards of education with student privacy responsibilities and 7 of these became law; 11 bills gave this responsibility to district or county school boards and 1 became law; 28 gave privacy or security responsibilities to local education agencies (LEAs - school districts) and 9 became law.
- In 2015, 35 bills charged state boards of education with student privacy roles and 5 of these became law; 23 bills tasked local school boards with the responsibility and 7 of these became law; and 62 gave privacy or security responsibilities to LEAs and 9 became law.
- In 2016, 13 bills charged state boards of education with student privacy responsibilities and 4 of these became law; 10 bills gave this responsibility to school or count boards and 4 became law; and 44 bill described privacy or security responsibilities for LEAs and 5 became law. [26]

At this point, it seems that states see shared responsibilities for privacy and security across all three levels of school governance structures. Many states also addressed concerns about the capacity and resource needs of school districts in managing the issues around student privacy, especially with respect to staff training and explicit policies such as those for contracts with service providers.

Over the last three years there has been increased attention across all states on the roles and responsibilities of ed tech companies. In 2014 California passed the first law explicitly targeting online providers in its Student Online Personal Information Protection Act (SOPIPA). Initially other states modeled their legislation after SOPIPA but increasingly states have expanded and

[25] DQC, 2015, p. 2
[26] Data Quality Campaign, 2014, 2015 and 2016.

strengthened its requirements, especially with respect to the requirements on ed tech companies.[27]

During 2015, Delaware enacted the Student Data Privacy Protection Act which "prohibits education technology service providers, primarily operators of Internet websites," cloud services and mobile services used for K-12 schooling purposes from selling data and using data and amassing a profile of students for non-educational services.[28] In addition to Delaware, Connecticut and Colorado introduced legislation in 2016 that went above and beyond common ambiguous requirements commonly listed in legislation.

Connecticut enacted the Student Data Privacy Act of 2016, which focuses on providing requirements to all contracts entered through a local or regional board of education and that all contractors are required to "implement and maintain security procedures and practices to protect"[29] student data. Connecticut's law requires certain security measures that Internet operators must follow, including notifying the public education entity during the case of a material breach, and it also requires edtech companies and Internet operators to make available the type of data collected, why it is being collected, how it is used and with whom the data is shared. On top of these requirements, Connecticut established a "task force" that is responsible for studying issues strictly related to student data privacy.

Colorado's Student Data Transparency and Security Act of 2016 is arguably the most effective legislation on student data privacy post-California, leaving no gaps or questions about the responsibilities of the contractors and third-party service providers, as well as the responsibilities of institutions, such as state departments and local education agencies. Colorado's strict student data privacy act aimed to heal the trust lacking between parents and school officials. It was passed unanimously by both the Democrat-controlled House and the Republican-dominated Senate. The act "defines what data can be collected, who can collect, for what purposes, how it gets held, and how it gets protected and what's done with that information"[30] in response to concerns that students' sensitive information is being collected and "potentially outsourced to third parties with no guardrails."[31] It tasks the Colorado Department of Education with creating transparency in regards to contracts with edtech software companies (including the type of student data being collected, how it is being collected, how it will be used, etc.) and requires the contractor to provide local education agencies (LEAs) and parents with this same transparency. The act prevents "educational software and app makers from collecting any

[27] For more information on SOPIPA, see: Dylan Peterson, "Edtech and Student Privacy: California Law as a Model," *Berkeley Technology Law Journal* 31:2 (2016): 961-995.

[28] Senator Sokola et al. (2015). Student Data Privacy Protection Act, Senate Bill No. 79. *Delaware State Senate, 148th General Assembly.* Retrieved from: https://test.legis.delaware.gov/json/BillDetail/GetHtmlDocument?fileAttachmentId=49631

[29] Connecticut General Assembly. (2016). Student Data Privacy Act of 2016, Public Act 16-189. *Connecticut General Assembly.* Retrieved from: https://www.cga.ct.gov/2016/ACT/pa/2016PA-00189-R00HB-05469-PA.htm

[30] Schrader, M. (2016). "Student privacy bill flies through Colorado Statehouse with unanimous support," *Colorado Springs Gazette*. http://gazette.com/student-privacy-bill-flies-through-colorado-statehouse-with-unanimous-support/article/1575565

[31] Bunch, J. (2016, April 14). *How computer privacy in Colorado high schools could be changing.* Denver Post. http://www.denverpost.com/2016/04/14/how-computer-privacy-in-colorado-high-schools-could-be-changing/

data that can be linked directly back to an individual student"[32] and requires them to notify LEAs in the event of a material breach of data and either investigate the issue or strengthen the security measures within a restricted time frame. The Colorado Act also require software companies in contracts with schools to "destroy" student information, if requested, rather than "delete" student information (which only means to "sever the address").[33]

Colorado's Student Data Transparency and Security Act has provided a new foundation for other states to pass strict legislation in regards to student data privacy.

[32] Bunch, J. (2016, May 5). *Colorado student data privacy bill on its way to becoming law.* The Denver Post. http://www.denverpost.com/2016/05/05/colorado-student-data-privacy-bill-on-its-way-to-becoming-law/
[33] Bunch, April 2016