Big Data in the Education Arena:
21st Century Student Sorting and Tracking[1]

Priscilla M. Regan
George Mason University

Jolene Jesse
National Science Foundation

Elsa Talat Khwaja
George Mason University

Prepared for Presentation at the Surveillance Studies Network
Conference, April 2016, Barcelona, Spain

Student tracking in the 1950's, especially in heterogeneous societies such as the United States, resulted in classrooms that were often divided by race, ethnicity, gender and class. Such tracking was glaringly obvious to parents, students, teachers and administrators – and thus the implications and wisdom of tracking became subjects of policy and social debate. In contrast, the student tracking that appears to be occurring in 2016 is hidden from view as it takes place behind computer screens in the varied educational programs that different students are exposed to – based on how big data applications have evaluated their likely learning profile.

Education is one of the arenas in which big data applications are presently being aggressively marketed, not only at the college level but perhaps even more so at the elementary and secondary level. As all countries recognize the importance of competing in the global environment and as the world becomes more of a global village as a result of economic and social activities facilitated by the Internet, countries around the globe are directing attention and resources on improving educational achievement especially at the primary and secondary levels. With the concomitant increase in the costs of providing education and concerns about financial responsibility, heightened consideration of accountability and results, elevated awareness of the range of teacher skills and student learning styles and needs, more focus is being placed on the promises offered by online software and educational technology. Information technology companies recognize the huge market offered by K-12 education and are aggressively developing and marketing their products. Most of these companies are large international ones based in the United States, such as Google and Microsoft, but a range of new startup companies, still largely based in the United States, now populate the market.

The benefits of big data applications include more sophisticated analyses of student learning and testing, more personalized learning, more effective delivery of educational materials, improved assessment, and more responsiveness to student needs. On the downside big data applications and products raise the possibility of discrimination as a result of profiling and tracking of students, as well as uses of student information for a wider range of purposes.

This paper seeks to first analyze the big data educational market – who are the key players; are they primarily country-specific firms or global firms; how are the benefits and downsides of big data applications being framed in marketing materials; what types of educational institutions, levels, and/or subjects are being targeted. Second, the paper will explore whether educators and/or civil society groups are responding to big data educational innovations – what discourse has resulted; what themes are being voiced and by whom; in what venues are discussions taking place. Third, the paper is particularly interested in discussions about whether and how categories such as race, gender, ethnicity, and class, as well as their intersections, are affected by big data applications in education – and what the implications of this are, particularly for children in their early educational years where opportunities for future learning may be critically shaped. Finally, the paper will close with a comparison of the policy and social debates in the 1950s about educational tracking and those of the 20'teens.

Big Data and Big Business in Education

Over the last twenty years, technology has become ubiquitous in classrooms at all levels, especially at the K-12 level. As noted above, this is explained by a number of factors including the focus on STEM education, the general social trend to a more technologically sophisticated society and the need to prepare upcoming generations to compete in that environment, and the fact that computer-assisted learning might entice students into engaging with material, which might normally have appeared less attractive to students. But other factors are also at work here, especially the pressure for student achievement, teaching effectiveness, controlling the school budget and also the interests of the technology companies in this seeming lucrative market. Additionally venture capitalists see the growth potential in the education market and are investing in ed tech start-ups – and finally, large wealthy foundations, such as the Bill and Melinda Gates Foundation, believe that technology offers many tools for improving the educational experience for children. The confluence of the importance of education achievement and effectiveness, the reality of the digital environment which students inhabit more generally, tighter educational budgets, and the profit interests of technology companies create an environment in which schools and departments of education are under pressure to adopt technology for a range of activities.

Educational technology is shaped by the changes in the education space itself; everything from financial constraints on schools and school boards, new demands for accountability and outcome measures, innovations in teaching and learning, and

new laws and requirements have driven the market for new technologies and the data generated by them to inform decision making around teaching, learning and policy. The education technology sector is booming, with more than $1.8 billion in venture capital currently being invested in the estimated $8 billion market for education technology software. This has meant a surge in new startups, increasing attention paid by large and established technology companies, and increased competition for domain space. During and in the wake of the InBloom saga, major companies like McGraw Hill Education, Pearson, and even News Corp had developed data tracking software for education.[2] Facebook, Google, and Microsoft have also acquired or are developing their own education software subsidiaries.

Much of the discussion about big data in education journals and newsletters reports on new initiatives conducted by educational firms, the promises of big data, and the positive effects on student learning and achievement. For example, Darrel West in a Brookings report presents several potential benefits of big data including insights regarding student performance and approaches to learning, effectiveness of techniques, evaluation of student actions, and predictive and diagnostic assessments. He also notes several barriers complicating the achievement of these benefits including the need for data sharing networks, similar data formats, and balancing vital student privacy and confidentiality with access to data for research purposes but cautions that "Using privacy arguments to stop research that helps students is counter-productive."[3]

In order to provide a concrete context for understanding how big data innovations raise ethical concerns, the following section provides an overview of the controversy surrounding InBloom in New York State.

InBloom: Controversy leads to legislation and bankruptcy

In the fall of 2013 twelve parents filed a lawsuit to stop an agreement between the State of New York and InBloom, a nonprofit corporation started by the Council of Chief State School Officers and underwritten by a $100 million grant from the Bill and Melinda Gates Foundation and the Carnegie Corporation of New York. At the time of the lawsuit, InBloom had commitments from nine states to adopt its cloud service, although only New York, Louisiana and Colorado had actually signed contracts and were undertaking pilot efforts to upload data with the non-profit. By October 2013, New York State had already uploaded 90 percent of the data from 2.7

---

[2] Ibid.
[3] Darrell M. West, "Big Data for Education: Data mining, Data Analytics, and Web Dashboards," *Governance Studies at Brookings* (September 2012).
http://www.brookings.edu/~/media/research/files/papers/2012/9/04-education-technology-west/04-education-technology-west.pdf

million public and charter school students into the system.[4] Education technology vendors also liked InBloom and were signing on to the service.

InBloom was supposed to be a data aggregator, meant to serve as a repository for the streams of data being generated by multiple education technology sources. InBloom would allow the data gathered from disparate educational software programs and apps to be uploaded into a cloud repository, translated into a common language, and made accessible through a dashboard by teachers, school administrators, school boards, and state departments of education, along with other "third parties". Users could then track individual students' progress through various educational stages, and teachers and others could intervene or "personalize" the learning experiences of individual students as they either struggled with or needed more challenge from the curriculum.[5]

In February 2014, the parents' lawsuit was dismissed, but by that point the New York State Legislature had put provisions in the state budget restricting the State Department of Education from undertaking any contracts with third party data aggregators. InBloom closed its doors in April 2014 after school districts in Louisiana and Colorado followed New York State's lead and pulled out of pilots involving the data repository.[6] What ultimately led to InBloom's demise was a cacophony of voices from many sides concerned about privacy, parental consent and access to the aggregated data. InBloom's software had included some 400 "optional fields" that schools could choose to fill in and that included some fairly sensitive information such as disability status, social security numbers, family relationships, reasons for enrollment changes, and disciplinary actions.

Parents and privacy advocates balked at what they saw as intrusive data gathering that seemed like surveillance. Questions were raised about who could and would access the data, especially data like disciplinary actions, with subjective terms like "'perpetrator,' 'victim,' and 'principal watch list,'" as well as the potential for such data to be used to "stratify or channel children."[7] Parents were particularly incensed that InBloom would not allow any opting out of the data collection. Teachers and other education professionals were concerned about state-level officials having access to student-level data, and about the potential use of

---

[4] Sanger, Natasha, October 5, 2013, "Deciding Who Sees Students' Data," The New York Times, http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html; accessed February 16, 2016.
[5] Sanger, Natasha, October 5, 2013, "Deciding Who Sees Students' Data," The New York Times, http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html; accessed February 16, 2016.
[6] Sanger, Natasha, April 21, 2014, "InBloom Student Data Repository to Close," The New York Times Bit Blog, http://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close/?_r=0, accessed March 8, 2016.
[7] Sanger, Natasha, October 5, 2013, "Deciding Who Sees Students' Data," The New York Times, http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html; accessed February 16, 2016.

sometimes dubious measures to assess the effectiveness of teachers in the classroom.

InBloom has insisted its efforts were misunderstood. As a data repository, InBloom officials maintained they were not controlling or using data, simply storing it for schools and school districts to have easier access across the large number of data platforms, software, and apps. In other words, they were to be a middleman between software vendors and school districts, with the districts controlling their own data.[8] InBloom was not alone in the data aggregation space; there are a number of data aggregators who are currently doing exactly what InBloom had promised to do, including Pearson (PowerSchool student information system) and Clever, based in San Francisco. Pearson and Clever both house data on 13 million school children and 15,000 school districts respectively.

However, InBloom got caught in the middle of the national debate about the future of education, and privacy became the issue that united the opposition and proved convincing to legislators that a limit had been reached. It didn't help that InBloom fought all efforts to allow parents to opt out of the service, and that the New York State Department of Education refused to listen to public concerns over security and access to the data. The controversy ballooned into a large scale lack of trust in InBloom and widespread perceptions that InBloom and the State were arrogant and insensitive.[9] Critics justifiably pointed out that InBloom and the NY State Department of Education hadn't fully assessed risks and liabilities surrounding both privacy and data security.

Rather than kill the education software sector, the demise of InBloom simply provided more space for other companies to come in and fill the void.[10]
At the same time, the policy issues that emerged from the fall of InBloom are increasingly leading to new discussions about privacy undertaken in new and existing arenas and with emerging actors in the policy space. Educational technology and particularly big data raise issues about the privacy and security of student data, the role of traditional educational actors – teachers, parents, school administrators, school boards, state departments of education, and national departments of education – as well as the role of new educational actors,

---

[8] Herold, Benjamin, April 21, 2014, InBloom to Shut Down Amid Growing Data-Privacy Concerns, Education week,
http://blogs.edweek.org/edweek/DigitalEducation/2014/04/inbloom_to_shut_down_amid_growing_data_privacy_concerns.html; accessed March 8, 2016.
[9] Bogle, Ariel, April 24, 2014, "What the Failure of InBloom Means for the Student-Data Industry," Slate Future Tense Blog,
http://www.slate.com/blogs/future_tense/2014/04/24/what_the_failure_of_inbloom_means_for_the_student_data_industry.html; accessed March 8, 2016.
[10] Bogle, Ariel, April 24, 2014, "What the Failure of InBloom Means for the Student-Data Industry," Slate Future Tense Blog,
http://www.slate.com/blogs/future_tense/2014/04/24/what_the_failure_of_inbloom_means_for_the_student_data_industry.html; accessed March 8, 2016.

particularly online and software education technology firms. The next section provides an overview of the major ethical issues that are emerging.

Ethical Policy Concerns about use of Big Data in Education

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. We can identify six concerns traditionally associated with privacy that are challenged by big data generally and in the context of education.

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.

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 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."[11]

Moreover, enhancements in digital storage capacity combined with improvements in computational power and developments of more sophisticated algorithms for analyzing data have enabled organizations to probe and dissect datasets in ways unimagined even twenty years ago. As Rubinstein similarly points out big data make possible the extraction of new, potentially useful information from data – this "newly discovered information is not only unintuitive and unpredictable, but also results from a fairly opaque process."[12] The entire enterprise of big data challenges

---

[11] 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

[12] Ira S. Rubinstein, "Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law* (2013) 3(2): 74-87. Available at: http://idpl.oxfordjournals.org/content/early/2013/01/24/idpl.ips036.full.pdf+html (pp.1-14), p.3

all previous ideas about how to limit data collection about individuals and how to involve the individual in the process of data collection and subsequent uses so that the individual could exercise some meaningful control.

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 providing information and controls about the uses of big data are at least one step removed from the data subject.

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.[13] 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.[14] In effect few characteristics are actually needed to uniquely identify an individual, making it very difficult to anonymize databases by removing some characteristics, because the bundle of characteristics remaining will likely prove sufficient to identify individuals once a database is merged with other databases and searched using sophisticated algorithms. For example, Latanya Sweeney and colleagues identified the names of volunteer participants in the de-identified public, Personal Genome Project by linking the Project's profiles to public records and data mining the results.[15]

---

[13] 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/
[14] 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
[15] 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

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, anonymized data – but this tradition is being challenged in the era of big data.

Computer scientists and privacy policy experts and advocates continue to press for better techniques for anonymizing data, for example by using only 3 digits of one's ZIP code or redacting year of birth or day of month. However, as databases become larger and more integrated these attempts increasingly prove to be ineffective. 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."[16] This is a conclusion that is becoming more widely shared as various big data projects by companies such as Netflix, AOL and Google reveal that individuals can indeed be identified in studies that were using supposedly anonymous data. And there is increasing recognition that data can either be useful or protective of privacy, but not both. As a biomedical researcher notes: "I can't anonymize your genome without wiping out the information that I need to analyze."[17] 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 – and that big data require to be even more powerful. A key element of this surveillance is what is now being referred to as the "internet of things," where all our smart devices pick up and transmit detailed information. Big data not only entails more monitoring of activities and extraction of data about those activities, but also involve analysis of those activities to determine likely future activities. This more sophisticated prediction that is built into many big data analytics transforms surveillance into a more omniscient phenomenon.

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 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 are cross-matched with more traditional information about the student as well as new information from various devices (such as how much a

---

[16] Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," 57 *UCLA Law Review* 1701-1777, 1704 (2010).
[17] 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

student moves throughout the day or how much time a student spends on social networking sites) – and all of this is 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 raises 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 – to buy certain items, try certain routes or restaurants – and in a certain way challenge the self as defined throughout much of Western philosophy. Some have expressed this concern as about social fragmentation into "filter bubbles," where individuals are subject to feedback loops that limit individuals' sense of their options.[18] 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.[19] Tene and Polonetsky point to the dangers of predictive analysis including the perpetuation of old prejudices and the accentuation of social stratification.[20]

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.[21] 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

---

[18] Eli Pariser, *The Filter Bubble: How the new Personalized Web is Changing What We Read and How We Think* New York: Penguin Books, 2011.

[19] 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.

[20] 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

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

"big data helps companies find a reasonable proxy for race."[22] 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. As Jonas Lerman points out: "The big data revolution may create new forms of inequality and subordination, and thus raise broad democracy concerns."[23] 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.[24]

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,[25] the property rhetoric and rationales have become part of the policy discussion about big data, as they had been in earlier iterations of debates about privacy policy. As one moves further from either submitting personal information to one organization or clicks "I agree" on a website, any ownership in that information arguably fades. And if that information becomes part of a dataset that is then reused or reconfigured or combined with another or sold to another organization, the claim of personal ownership in that information diminishes even more. 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.

Policy Discussions – Responses to Ed Tech and Big Data

In the United States, the issue of privacy and student data emerged as a topic in state level policymaking due to the Snowden revelations and concerns about data surveillance generally, publicity surrounding data breaches at retailers such as Target, the increased use of educational technology for a number of functions – administrative systems, classroom instruction, homework, student collaborations,

---

[22] 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.

[23] Jonas Lerman, "Big Data and Its Exclusions," *Stanford Law Review Online* (Sept. 3, 2013) 66: 55-63, 60.

[24] 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

[25] 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).

and incorporation of social networking, and the InBloom controversy. In 2013, Joel Reidenberg directed a study on cloud computing in public schools, which found that school districts were not addressing privacy concerns in a uniform or informed manner when they transfer student information to cloud computing service providers.[26] Based on their detailed investigation into a sample of schools, they concluded that "cloud services are poorly understood, non-transparent and weakly governed"[27] and "an overwhelming majority of cloud services do not address parental notice, consent, or access to student information."[28]

In 2014, [29] 110 student data privacy bills were introduced in 36 states with 21 states passing 24 such bills into law. The latter half of 2014 saw a shift in policy discussions from concern with data in state systems to the privacy implications of student data collected, held and analyzed by third party service providers following the controversies and press attention from InBloom's activities in New York and Colorado. California passed the first law explicitly targeting online providers in its Student Online Personal Information Protection Act (SOPIPA). State legislative interest increased in 2015 with 182 student privacy bills introduced in 46 states and 15 states passed 28 new student privacy laws. Bills in 25 states were modeled after SOPIPA and in 31 states bills articulated requirements for service providers. 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.

The Data Quality Campaign identified two overlapping approaches in these state bills. First, the *prohibitive approach*, which restricted or prevented the collection of certain types of data (e.g., biometric) or certain uses of data (e.g., predictive analytics), was adopted in 79 of 110 bills in 2014 and 125 of 182 bills in 2015. Second, the *governance approach*, which established procedures (e.g., audits and inventories), roles and responsibilities to ensure appropriate student data practices, was found in 52 of 110 bills in 2014 and 122 of 182 bills in 2015.[30] States are still sorting out 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; 28 bills gave this responsibility

---

[26] Joel Reidenberg, N. Cameron Russell, Jordan Kovnot, Thomas B. Norton, Ryan Cloutier, and Daniela Alvardado, "Privacy and Cloud Computing in Public Schools," Center on Law and Information Policy, Fordham Law School (Dec. 13, 2013) available at:
http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1001&context=clip
[27] Ibid, p. 6
[28] Ibid, p. 7.
[29] Material regarding state laws is derived from two publications from the Data Quality Campaign : "State Student Data Privacy Legislation: What Happened in 2014, and What is Next?" (August 2014) available at: http://dataqualitycampaign.org/wp-content/uploads/2014/09/DQC-Data-Privacy-whats-next-Sept22.pdf
and "State Student Data Privacy Legislation: What Happened in 2015, and What is Next?" (September 2015) available at: http://dataqualitycampaign.org/wp-content/uploads/2015/09/DQC-Student-Data-Laws-2015-Sept23.pdf
[30] Data Quality Campaign, 2015

to school districts. In 2015, 35 bills charged state boards of education with student privacy roles and 5 of these became law; 63 bills gave this responsibility to school districts and 9 became law; and 23 bills tasked local school boards with the responsibility and 7 of these became law.[31]

There is variation across school districts and individual schools in what kind of educational technology is being used, for what purposes, under what guidelines or policies, and with what kind of oversight. As Joel Reidenberg's 2013 survey of cloud computing practices found, variation in practices and policies is quite wide – and a preliminary search for emerging big data applications at the school district and school level in the US reveals the same varied pattern but with many schools or school districts not having yet addressed this issue. The US Department of Education's Privacy Technical Assistance Center outlined several best practices for schools and school districts, including conducting an inventory of online educational services being used, establishing policies for approving proposed online educational services, using written contracts or agreements, and being transparent with students and parents. Perhaps most revealing of activities at the school and district level were the cautions about teachers' use of "Click-Wrap" software which enters the end-user (which in this case would be the school or district) into a signed contract (a contract that the software company may update or amend without notice) and about use of "free" online services that may pose the same risk to student privacy and security.[32]

A preliminary review of ed tech company websites[33] reveals 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 (see below). Most privacy policies, once found, are written in somewhat accessible 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

---

[31] Data Quality Campaign, 2014 and 2015.

[32] 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, pp. 7-10.

[33] 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.

change the policy at any time," and that the information "may be subject to an entirely new privacy policy, if the company is sold – a common fate for a start-up."[34]

So while companies have taken advantage of the opening up of the data aggregation space in light of InBloom, they 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.

At the same time that the education technology sector has exploded, the student privacy advocacy space has also welcomed new actors, both in favor of the increased use of big data in education (pro-big data) and for more attention paid to privacy and accountability (pro-privacy).

Those advocating for the development of education technologies tend to focus on the benefits of using technology in offering teachers, schools and education policy makers the kind of evidence that would lead to more success in the classroom. These groups, including the Data Quality Campaign (DQC), the Future of Privacy Forum (FPF), the Consortium for School Networking (CoSN), the Student Privacy Resource Center (FERPASherpa), the Software and Information Industry Association (SIIA), and the iKeep Safe Coalition, among others, are largely funded by large technology corporations and their foundation arms. Their websites and informational brochures tend to focus on the benefits of using technology and data in the classroom, along with information about the various privacy laws and current student privacy protections. Their audience tends to be school leaders and the technology companies themselves, with secondary attention paid to teachers and parents.

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.[35] 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."[36] 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

---

[34] 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

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

[36] 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.

training modules and awards for state and local officials who "have embraced a culture of data in service of students."[37]

Some question the efficacy of privacy pledges and certification, even as the President and others have embraced the movement. 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.[38] 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.[39]

On the pro-privacy side, parent and teacher groups have begun to coalesce around the privacy issues and existing privacy organizations have focused more and more on student privacy in the educational context. The Electronic Frontier Foundation (EFF) and the Electronic Privacy Information Center (EPIC), for example, are well-established in the privacy arena and have increasingly advocated for student rights.[40] The InBloom controversy spurred more parent and teacher organizations to hone in on privacy issues including the Parent Coalition for Student Privacy, an advocacy group started by Leonie Haimson, a parent advocate in New York City, and Rachael Strickland of Colorado. The National Parent-Teacher Association (PTA) also has weighed in with a statement on student privacy.[41]

These organizations have been particularly effective at articulating the objections to the types of data aggregators exemplified by InBloom, including the threats to student privacy through the weakening of FERPA, data sharing practices among school districts and states, the development of longitudinal data tracking systems,

---

[37] 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.
[38] 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.
[39] 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
[40] Privacy statements and other information from these organizations are available online: EFF Tips for Parents https://www.eff.org/issues/student-privacy/tips; EPIC Student Privacy Bill of Rights https://epic.org/privacy/student/bill-of-rights.html;
[41] National PTA Position Statement on Student Data Privacy and Security http://www.pta.org/about/content.cfm?ItemNumber=4614, accessed March 9, 2016; Parent Coalition for Student Privacy http://www.studentprivacymatters.org, accessed March 9, 2016.

and the push for continuously quantifying students for personalize learning. Of particular concern is the sharing of data with for-profit data-mining vendors and other third party commercial concerns who might then market products directly to students, or the theft of student data by hackers.[42] These organizations generally do not accept money from the large technology companies. Their target audiences tend to be teachers, parents and students. Instead of pledges or seals, the focus from these organizations is on information, with "student bills of rights" or legislative action campaigns the major advocacy activities.

In addition to advocacy groups, technology associations, and parent groups, teachers unions have also participated in policy discussions about big data and education. The United Federation of Teachers (UFT) for example in December 2013 urged New York State to end its contract with InBloom, as reflected in this statement from its Vice President before the state Assembly Education Committee: "The UFT is not opposed to gathering data on public school students; in fact, it's a valuable tool [but the initiative with inBloom] is about releasing sensitive, student-identifying data points in 400 categories ... and, ultimately, share[ing] some or all of that information with private companies developing education software. How can we possibly countenance that?"[43] This position had been foreshadowed in the UFT's May 2013 press release: "Any potential inBloom has to improve and personalize learning is being overshadowed by a growing lack of public trust in its early communications and operations, and genuine concerns about the security, privacy, sharing and exploitation of data."[44]

At this point, it is difficult to tell whether the pro-big data or pro-privacy protection forces have the upper hand in the debate, although both do seem to have an influence on the discussions. New state laws, for example, often incorporate concerns about data access, limited data collection, data use, security, transparency, and accountability advocated by the pro-privacy protection groups. The federal government, especially the executive branch (White House, the Office of Science and Technology Policy, and the Department of Education) have more often embraced the increased use of technology and data to make more informed decisions about how to use big data to increase student success with a more flexible view on data privacy.

Policy discussions about the ethical issues raised by big data and education are likely to continue. At this point in the debate, we can identify a number of themes and trends. The first is that the current focus is largely on:
- the security of the data
- deidentification of student data for analytical purposes

---

[42] Kharif, Olga, May 1, 2014, "Privacy Fears over Student Data Tracking Lead to InBloom's Shutdown," Bloomberg Business, http://www.bloomberg.com/bw/articles/2014-05-01/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking; accessed March 8, 2016.

[43] Micha Landau, "Incensed Over InBloom," (Dec 19, 2013), available at: http://www.uft.org/public-ed-under-attack/incensed-over-inbloom

[44] See more at: http://www.aft.org/press-release/aft-statement-privacy-and-security-concerns-about-inbloom-and-other-data#sthash.lPbRWN9P.dpuf

- prohibitions on targeted advertising using student data
- ownership of information, with general agreement that ownership should remain with the local school district and
- transparency regarding online practices, including inventories of online and ed tech programs.

The second trend is that the discourse of these ethical issues is shaped primarily by the current legal framework and standard fair information principles. Government policy documents in particular begin their analyses with questions about whether and how existing privacy statutes, incorporating the FIPPS framework, apply – and how they might be amended if they do not provide adequate coverage. This fairly quickly locks the policy discussion into pre-existing categories and may ignore other approaches and preclude a fresh look at the issues.

The result of these first two trends is that the issue of profiling of students and potential discriminatory effects resulting from big data analytics has not yet been incorporated directly into these evolving policy discussions. Discrimination and social/panoptic sorting have been important topics of information privacy and data protection debates but cast more under the rubric of "surveillance" rather than privacy – and not therefore directly addressed as such in privacy statutes. Additionally concerns about discrimination have also been framed more as civil rights or human rights issues. In the US, under what the Data Quality Campaign calls the "prohibitive approach," there were proposals at the state level during 2015 to ban "predictive analytics" or to prohibit the collection of certain arguably more sensitive types of data, such as social or emotional learning data and biometric data. These types of prohibitions more directly confront the possibilities that big data could be used in discriminatory ways.

The final section of the paper discusses the discriminatory potential of big data applications in the context of the larger and far longer debate about student tracking.

Student Tracking – More Sophisticated Panoptic Sorting with Big Data

One education commentator noted that "The most enduring feature of the American education system is its character as a sorting machine."[45] American tracking of students is generally more subtle and obscured than educational tracking in Europe where students have long been sorted, based generally on test scores, by the time they start their secondary education into vocational and technical schools or into college-preparatory schools. Instead in the US, tracking is incorporated into schools where different classes may be labeled "gifted and talented," or "advanced

---

[45] Marc Tucker, "Student Tracking vs Academic Pathways: Different...or the Same?" *Education Week* (October 15, 2015). Available at:
http://blogs.edweek.org/edweek/top_performers/2015/10/tracking_vs_pathways_differentor_the_same.html

placement" or "honors" classes in contrast to the "general" track or "special needs" classes. In elementary schools, tracking may occur under the guise of colors or birds. Tracking may also occur within a class itself where "ability groups" are given more or less challenging material based on the teacher's perception of their ability and their test scores. Regardless of the labels used tracking continues to be used in some form throughout the K-12 level in most American schools – and has been controversial for more than a century.[46]

The debate about tracking began in the late 1800's when more students were going on to public high schools rather than finishing their formal education at the end of elementary school. An influential 1893 report of what was termed the "Committee of Ten," appointed by the National Education Association (NEA), chaired by the then President of Harvard, concluded that all public high school students, consistent with the principle of equal opportunity for all and the significant role that education could play in achieving equal opportunity, should take a college prep curriculum. In 1918 a second NEA committee, the Commission on the Reorganization of Secondary Education, issued another report recommending more differentiated high school programs to take into account the variety of abilities, goals and financial means of the more diverse student population resulting from immigration. As a result, by the mid-1920smost urban high schools offered four high school tracks: college prep, commercial (office work, mainly aimed at female students), vocational (home economics and industrial arts), and general.[47]

During the 1960's, the racial and class effects of student tracking received much attention and criticism – and these concerns persisted through the 20th century with the consensus being that "tracking has minimal effects on learning outcomes and profound negative effects on equity outcomes."[48] As recently as 2014 the US Department of Education and critics of tracking have expressed concern that "tracking perpetuates a modern system of segregation that favors white students and keeps students of color, many of them black, from long-term equal achievement."[49] A number of factors account for the persistence of tracking

---

[46] See generally: Tom Loveless, "The Resurgence of Ability Grouping and Persistence of Tracking," (Part II of the 2013 Brown Center Report on American Education," *Brookings Report* (March 18, 2013), available at: http://www.brookings.edu/research/reports/2013/03/18-tracking-ability-grouping-loveless and Carol Corbett Burris and Delia T. Garrity, *Detracking for Excellence and Equity,"* see especially Chapter 2 "What Tracking is and How to Start Dismantling It," available at: http://www.ascd.org/publications/books/108013/chapters/What-Tracking-Is-and-How-to-Start-Dismantling-It.aspx

[47] Jeffrey Mirel, "The Traditional High School: Historical Debates Over its Nature and Function," *Education Next* (Winter 2006)/Vol 6, No. 1, pp. 14-21. Available at: http://eric.ed.gov/?id=EJ763310

[48] John Hattie's results of a meta-analysis quoted in Valerie Strauss, "The Bottom Line on Student Tracking," *The Washington Post* (June 10, 2013). Available at: https://www.washingtonpost.com/news/answer-sheet/wp/2013/06/10/the-bottom-line-on-student-tracking/

[49] Sonali Kohli, "Modern-Day Segregation in Public Schools," *The Atlantic* (Nov. 18, 2014). Available at: http://www.theatlantic.com/education/archive/2014/11/modern-day-segregation-in-public-schools/382846/

including a genuine concern about student learning and effectiveness but also the reality that some parents are better able to "game" the system to the advantage of their children. Additionally the No Child Left Behind Act of 2001 and its focus on test scores and lower achieving students resulted in more "targeted instruction…and an increase in de facto tracking in younger grades."[50] As one parent noted "You see kids entering the building through the same door…But the second door is racially stratified."[51]

The debate about student tracking has continued into the 21st century with the focus increasingly being placed on the tracking that goes on "behind the computer screen" – the second door now more difficult for parents and even teachers to see. But as Tom Loveless points out in a Brookings report:

> The increased use of computer instruction in elementary classrooms cannot help but make teachers more comfortable with students in the same classroom studying different materials and progressing at different rates through curriculum. The term "differential instruction," while ambiguous in practice, might make grouping students by prior achievement or skill level an acceptable strategy for educators who recoil from "ability grouping."[52]

Big data applications in education signal yet another fundamental change in the dynamics of sorting students. The actions of today's "digital student"[53] are monitored and tracked in ways inconceivable in earlier times – and with the results of more fine-grained tracking, less transparency, and persistent record-keeping from pre-school through college and possibly beyond. Our review of the ed tech companies offerings and marketing materials indicates that these companies are amassing quite detailed information on student demographic characteristics in their databases (including not just traditional location and family information but: school lunch eligibility, emergency contact information, parent and guardian information, health profiles, disciplinary records, counseling referrals, etc.) as well as detailed information on student learning records (including not just test scores and grades but also individual learning and test-taking patterns, attention spans) – and all of this is analyzed with sophisticated algorithms resulting in new categorizations and groupings of students. Moreover, these records follow students throughout their educational careers. Whether these sortings replicate or serve as proxies for traditional discriminatory groups or create new ones may be something of an open question but one that is critical to purse.

---

[50] Ibid, remarks of Christina Theokas, research director for Education Trust, an advocacy group.
[51] Ibid
[52] Loveless, op cit.
[53] For review of legislative framework see: Center for Democracy and Technology, *Privacy and the Digital Student* (May 2015), available at: https://cdt.org/files/2015/06/Student-Privacy-White-Paper-v.-9_1.pdf

Pursuing data on this, however, appears difficult. Much of this information is proprietary and held tightly by the ed tech companies – in a way that teachers themselves do not understand. Legislative efforts to regulate or restrict the use of predictive analytics have not been successful, meeting as they do with opposition from the ed tech industry. Parents do not have full information about software packages and tests to ask questions – and efforts to require inventories of ed tech applications often result in more summary, and less useful, information.