Ethics and Information Technology (2019) 21:167–179
https://doi.org/10.1007/s10676-018-9492-2

ORIGINAL PAPER

Ethical challenges of edtech, big data and personalized learning: twenty-first century student sorting and tracking

Priscilla M. Regan[1] · **Jolene Jesse**[2]

Published online: 3 December 2018

Abstract
With the 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. One of the most heavily marketed, exciting and controversial applications of edtech involves the varied educational programs to which different students are exposed based on how big data applications have evaluated their likely learning profiles. Characterized most often as 'personalized learning,' these programs raise a number of ethical concerns especially when used at the K-12 level. This paper analyzes the range of these ethical concerns arguing that characterizing them under the general rubric of 'privacy' oversimplifies the concerns and makes it too easy for advocates to dismiss or minimize them. Six distinct ethical concerns are identified: information privacy; anonymity; surveillance; autonomy; non-discrimination; and ownership of information. Particular attention is paid to whether personalized learning programs raise concerns similar to those raised about educational tracking in the 1950s. The paper closes with discussion of three themes that are important to consider in ethical and policy discussions.

Keywords Privacy · Discrimination · Big data · Autonomy · Education technology · Personalized learning

The last 10 years have witnessed an explosion of new educational technologies (edtech), some touting amazing potential to reach the next generation with new learning methods that will teach not only content, be it history, mathematics or engineering, but also intra- and inter-personal competencies, such as resilience and teamwork. The edtech sector is actively marketing these learning tools, especially to elementary and secondary schools, although the efficacy of technology enhanced learning is still under investigation. Edtech applications have appeared at a political, policy, and commercial moment favorable to the capabilities and advantages offered. The increase in the federal, state and local costs of providing K-12 education and government and voter concerns about financial responsibility generate interest in new techniques that promise to improve efficiency of educational operations. Focus on student achievement and the rankings of US schools with those of other countries has led to heightened consideration of accountability and results. Elevated awareness of the range of teacher skills, as well as variations in student learning styles and needs, has drawn attention to the value of understanding unique characteristics of students and teachers. As a result, the K-12 school environment is conducive to the promises offered by online software and edtech. Edtech companies recognize the huge market offered by K-12 education—an arena that has a vast and renewable population base, but also a particularly vulnerable population involving minor children who experience a range of developmental milestones during the K-12 years.

This uptick in adoption of a variety of edtech applications at the K-12 level has also generated myriad policy debates, including proposed updates to existing federal laws and the introduction and adoption of numerous new state laws. Much of the policy debate is subsumed under the label of "privacy," although there are a range of ethical issues associated

✉ Priscilla M. Regan
pregan@gmu.edu

Jolene Jesse
jjesse@nsf.gov

[1] Schar School of Policy and Government, George Mason University, MSN 3F4, Fairfax, VA 22031, USA

[2] Division of Research on Learning in Formal and Informal Settings, National Science Foundation, 2415 Eisenhower Avenue, Alexandria, VA 22314, USA

with edtech applications that have not received the same amount of consideration as privacy, and some issues have been conflated with privacy. Privacy is certainly an issue, as the use of edtech entails collection of more, and more granular, information about students, teachers, and families, as well as administrative details regarding the functioning of educational institutions. Edtech applications enable sophisticated searching and analysis of collected information linking changes in the education arena to the larger debates about the challenges of big data generally. One of the most problematic aspects of edtech, and least addressed from a policy perspective however, involves the capability of edtech to deliver more personalized learning based on the needs and skill levels of individual students.

Personalized learning applications are currently among the most heavily-marketed, exciting and controversial applications of edtech. These applications involve evaluating students likely learning profiles on applications that use big data to categorize individual learning styles and then direct appropriate learning activities to those students. Known under several labels—personalized learning, student-centered learning, and adaptive learning—they are advocated by edtech companies and foundations, including the Bill and Melinda Gates Foundation and the Chan Zuckerberg Foundation. In 2016, 97% of school districts surveyed by the Education Week Research Center indicated they were investing in some form of personalized learning (Herold 2017). Although exactly what types of programs constitute personalized learning is not always clear and whether and how much these programs incorporate edtech is hard to determine, RAND in the third of its reports on personalized learning cautions that the evidence for the effectiveness of personalized learning is currently weak and needs more research in a range of school settings (Pane et al. 2017).

A critical ethical concern raised with personalized learning is whether such programs constitute tracking and sorting of students that might be considered discriminatory. The history of tracking in the United States is especially problematic, suggesting the need for caution when sorting children. Student tracking in the 1950s 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 2018 is hidden from the view of students, parents and even teachers as it takes place behind computer screens. The extent to which students might recognize they are being tracked through computer programs, and the impact that might have on learning outcomes is rarely discussed or researched. Similarly, the extent to which edtech software embeds subtle discrimination is also unclear, despite the current dialog about algorithmic bias.

This article seeks to first analyze the range of ethical issues raised by the increased use of edtech and big data in school systems throughout the United States—how these issues are framed; whether the major concerns are receiving the appropriate level of attention and analysis; and what policy implications there are around how issues are being presented. Second, the paper briefly explores policy responses to big data educational innovations—what discourse has resulted; and what policy trends are emerging. Third, the paper is particularly interested in personalized learning systems and whether and how they might incorporate categories such as race, gender, ethnicity, and class, as well as their intersections, and whether discussions about these systems mirror the concerns of the policy and social debates in the 1950s about educational tracking. Finally, the paper closes with some themes that are important to consider in ethical and policy discussions addressing personalized learning systems.

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, 12 parents concerned about the privacy of student records 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 million public and charter school students into the system (Singer 2013).

InBloom was supposed to be a data aggregator, designed to serve as a repository for the streams of data being generated by multiple edtech sources. InBloom would enable 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 (Singer 2013).

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 (Singer 2014). 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 (Bulger et al. 2017). 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 regarding disciplinary actions, with subjective terms like "perpetrator," "victim," and "principal watch list," as well as the potential for data to be used to "stratify or channel children" (Singer 2013). 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 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 substantial 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 (Herold 2014). InBloom was not alone in the data aggregation space; there are several 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 (Bogle 2014). 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.

The demise of InBloom, rather than halting interest in educational data aggregation, provided more space for other companies to come in and fill the void (Bogle 2014). At the same time, the policy issues that emerged from the fall of InBloom are increasingly leading to discussions about privacy in new and existing arenas and with emerging actors in the policy space. Edtech 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, particularly online and software education technology firms. InBloom's focus on aggregating the school-level data of elementary and secondary students was particularly susceptible to arguments about privacy and data control, leading to legal remedies and the rise of a number of new and proposed state laws to protect student privacy. 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 early discussion about edtech and big data in education journals and newsletters reported on new initiatives conducted by educational firms, the promises of edtech and big data, and the positive effects on student learning and achievement. Following the publicity around InBloom, there has been more discussion about ethical issues concerning the adoption of edtech in the K-12 environment, which has led to policy debates especially in state legislatures as these ethical issues deal with minor children, a perceived vulnerable population, and are centered on education, a contested space at all levels. As was the case with InBloom, subsequent policy and ethical discussions are most often framed in terms of "privacy." This is not particularly surprising both because privacy is viewed as a multi-faceted concept incorporating several distinct ethical concerns (Westin 1967; Solove 2008) and also because discussions about ethics and information technology in the United States have traditionally been categorized under the value of privacy (Regan 1995). This is equally true in the education sector with the federal Family Educational Rights and Privacy Act of 1974 and similar state laws framed in terms of privacy.

Although privacy is indeed a fundamental value potentially affected by edtech applications in a number of ways,

we argue that it is important to identify the specific "privacy" interests raised by edtech and to analyze each individually. Grouping all the ethical issues involved here under the general rubric of privacy makes it too easy for policymakers and the media to gloss over the complexity and breadth of concerns. A review of edtech articles from 2013 to 2017 in commercial and professional teacher-oriented publications revealed that discussion about ethical issues highlighted privacy issues framed almost exclusively in terms of protecting student information from inappropriate access or secondary uses and discussed in terms of compliance with standard fair information practices (Regan and Bailey 2018). A similar framing of issues occurs in the media including well-researched articles about edtech by Natasha Singer in the *New York Times*. This focus on privacy makes it easier for edtech advocates to minimize or simplify ethical concerns. Moreover, the general rubric of privacy allows policy makers to reduce the problem by emphasizing a broad definition of the issue for legislative "fixes" rather than adopting a more nuanced approach based on the complexity of the problem they are actually trying to solve. For example, Darrel West in a Brookings report presented 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 then captures these concerns by cautioning that "using privacy arguments to stop research that helps students is counter-productive" (West 2012). Similarly a 2016 report from the Center for Data Innovation, a research institute affiliated with the industry-oriented Information Technology and Innovation Foundation, listed "seven major obstacles to building data-driven education, including institutional resistance, hostility to using data in the classroom, a lack of effective tools, inadequate teacher training, flawed data infrastructure, systemic 'chicken or egg' challenges, and, *perhaps most significantly, privacy fears*" [emphasis added] (New 2016, p. 19).

In order to provide a more complete understanding of the ethical issues associated with edtech applications, we identify six concerns traditionally associated with privacy that are challenged by big data generally (Regan 2017) and particularly in the context of K-12 education. Each of these concerns is often categorized as "privacy," but as developed below, each is a distinct ethical concern, should be labelled as such, should be analyzed individually, and needs separate policy consideration and response. The six "privacy" concerns are: information privacy; anonymity; surveillance; autonomy; non-discrimination; and ownership of information.

The ***first*** concern raised by big data 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 most often been addressed by the fair information practice principles (FIPPs) generally summarized as notice, consent, choice and transparency. These principles are the basis of much privacy and data protection legislation around the world (Bennett 1992; Flaherty 1989).

Although many have questioned the effectiveness of the FIPPs approach especially in the United States where implementation relies on individual initiative (Gellman 1993; Schwartz 2000), there is almost universal agreement among privacy scholars and experts that the FIPPs approach will not be effective in the big data environment (Regan 2017, Barocas and Nissenbaum 2014; Ohm 2014). With big data there is more collection of information, by more parties, about more aspects of an individual's life, and with more granularity on the details of that life. Not only is there more information collected from more sources but much of the data collection takes place without the individual's awareness. 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 20 years ago. As Rubinstein 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" (2013, p. 3). The entire enterprise of big data challenges 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.

In the education arena, a National Academy of Education Workshop Summary distinguishes two types of big data. The first is "administrative data" including demographic, achievement and behavioral data collected by the schools, other government agencies, and contractors. Attendance records, test scores, transcripts, school lunch eligibility, and individualized education plans would fall into this category. The second category is "learning process data" including "continuous or near-continuous, fine-grained records, usually of digital interactions of student behaviors to illuminate learning processes" (2017, p. 3). Edtech applications are being employed for both types of data and both involve fairly detailed information about individual students and their families.

With respect to education and big data, issues of notice, choice, consent and transparency become even more

complicated than in other contexts because records of children and hence the concerns of parents come into play, and because the educational relationship is mandatory, not voluntary, thus obviating or restricting a realistic choice option. Additionally, edtech firms do not generally have a direct 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. In this environment, giving individuals some control over their information is untenable and control in effect passes to the school or teacher and the contracts that are negotiated with the edtech companies.

Two federal educational statutes follow the traditional FIPPS framework and give students and parents some rights with respect to notice, consent and transparency. First is the Family Educational Rights and Privacy Act of 1974 (FERPA), mentioned above, which requires schools to acquire consent before disclosing student information but allows a number of exceptions including to "organizations conducting certain studies for or on behalf of the school."[1] Arguably this might include edtech companies depending on the nature of the study. Additionally, collection and dissemination of information on students may be subject to the Children's Online Privacy Protection Act (COPPA) of 1998, and amended in 2013, affecting primarily private sector activities and enforced by the Federal Trade Commission. The application of these laws to big data in education is still unclear and was the topic of a December 2017 workshop cosponsored by the FTC and Department of Education. Most education and legal experts agree with Elana Zeide who concludes "FIPPS-based privacy protection is both ineffective and theoretically unsound in the education context" (2016, p. 107).

A ***second*** concern long associated with privacy is that individuals should be able to remain anonymous or obscure if they so choose. But with an ever-increasing number of social relationships and practices becoming data points, it becomes more difficult for individuals to remain unidentified or unidentifiable. Algorithmic searches of datasets now can rather quickly diminish what had been high transaction costs on finding meaningful information (Hartzog and Selinger 2013a, b). 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 obscured 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 (Sweeney 2000). In reality few characteristics are actually needed to identify a unique individual, making it almost impossible to anonymize databases by removing some characteristics as the remaining characteristics will likely prove sufficient to identify individuals once a database is merged with other databases and searched using sophisticated algorithms. For example, Sweeney et al. (2013) 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.

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 departments of education 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 assumption of anonymity of deidentified data is being challenged in the era of big data as it becomes easier to reidentify students. A National Academy of Education Panel Summary noted: "As more rich audio and video data are collected on group and individual work for assessment of social-emotional and soft skills [in K-12 classrooms], deidentifying data is untenable" (Bienkowski 2017, p. 1).

Two federal statutes address concerns about confidentiality in such data sets. The Protection of Pupil Rights Amendment (PPRA) of 1978 governs the administration to students of a survey, analysis, or evaluation that concerns one or more of eight protected areas, including: political beliefs; psychological problems; sex behavior or attitudes; anti-social or demeaning behavior; and religious beliefs. Secondly, the Education Sciences Reform Act of 2002 strengthens confidentiality requirements for student records especially with respect to the activities of the National Center for Education Statistics (NCES). Neither, however, addresses the possibilities for reidentification or culling of longitudinal datasets that is now possible with aggregation of multiple datasets and algorithmic searches.

A ***third*** concern that is often subsumed under the privacy rubric 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. 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" (2014, p. 38). Big data not only entails more monitoring of activities and extraction of data about those activities, but also involves analysis of those activities to determine likely future activities. This

[1] Available at: http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html.

more sophisticated prediction that is built into many big data analytics transforms tracking and surveillance into a more powerful tool that can be wielded in ways that have not yet been identified and understood.

Edtech enables more fine-grained and continuous observation of students through the collection of learning process data. Online testing and teaching programs monitor how long it takes students to answer a question or read a page—and often capture key strokes or patterns of reading or responding that might shed light on the thought processes of the student. These edtech applications also can follow where (home, school, computer lab) the student is working and what time of day—and may 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 can 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 in order, as the Department of Education's report points out "to maximize learning effectiveness and efficiency" (Bienkowski et al. 2012, p. 32)—raising a ***fourth*** ethical concern regarding autonomy.

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 or nudging 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 in terms of social fragmentation into "filter bubbles," where individuals are subject to feedback loops that limit individuals' sense of their options (Pariser 2011). 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 (Kerr and Earle 2013). Consequential predictions compromise autonomy, or individual agency, the least while preemptive predictions pose a greater threat to individual autonomy.

The ethical issues of nudging are often seen as problematic depending on the circumstances (Lichtenberg 2016; Kelly 2016; English 2016) but seem particularly problematic in education. Cass Sunstein acknowledges that "the ethical issues largely turn on whether nudges promote or instead undermine welfare, autonomy, and dignity…[and] a concern for personal agency often motivates the most plausible objections to nudges" (2015, p. 414). He tends to see nudges as generally helpful and defensible on ethical grounds as they often promote social welfare, provided that they incorporate transparency and accountability as safeguards. He cautions that the ethical danger with nudges occurs if they lead to manipulation instead of "steer people in particular directions but that also allow them to go their own way" (2015, p. 417). Although personalized learning systems may appear to be educative and in the student's best interest, the choice architecture of the prompts in these systems may be designed to entail more direction than suggestion. In this respect, by Sunstein's definition, personalized learning systems would appear to be manipulative to the extent that they attempt to "influence people in a way that does not sufficiently engage or appeal to their capacities for reflective and deliberate choice" (2015, p. 443).

Autonomy is related to a ***fifth*** concern associated with big data, and often folded under the privacy rubric in policy discussions, which involves traditional due process, 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 the information of others raises questions about treating individuals as individuals fairly, accurately, and in ways they can understand (Citron and Pasquale 2014). Tene and Polonetsky point to the dangers of predictive analysis including the perpetuation of old prejudices and the accentuation of social stratification (2013). This concern involves issues of profiling and discrimination.

Education systems and school districts, recognizing the importance of education for equal opportunity, have adopted policies and procedures to mitigate longstanding concerns about discrimination and to watch closely for subtle, as well as obvious, signs of discrimination. But with big data and edtech applications such subtle signs may be difficult to discern. For example, Ohm points out that "big data helps companies find a reasonable proxy for race" (2014, p. 101). Perhaps more troubling in education is the possibility that big data facilitates the creation of more refined, intersectional categories that might discriminate among students in 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." (2013, p. 60) 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 (Alarcon et al. 2014).

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 "own" the information? Although many privacy scholars question whether the property model provides a workable framework for talking about privacy (Cohen 2000; Schwartz 2004), 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 traditionally "owned" by the school or school district. The involvement of edtech companies has somewhat muddied the question of ownership—depending on how contracts with these firms are written. Heather Roberts-Mahoney et al. conclude that personalized learning "reconstructs the personal characteristics of students into the assets—private property—of database creators and education technology vendors" (2016, p. 13). One of the most problematic issues involves whether edtech companies should be able to use data generated by students' use of their software programs to improve those programs, raising questions about whether the companies are using students as test subjects for development and marketing of future edtech products. Elana Zeide refers to this as "beta" education where edtech companies and researchers "conduct what are essentially experiments on students when testing out different innovations" (2017, p. 516).

The above discussion helps to identify and distinguish the different ethical concerns that have been raised regarding edtech applications and to explain how these are related to the numerous interests that often are packaged as privacy (Westin 1967; Solove 2008). Although scholars, particularly in law review articles discussing potential biases posed by big data applications (Barocas and Selbst 2016; Citron and Pasquale 2014; Crawford and Schultz 2014) and in articles about nudging (discussed above), recognize discrimination and autonomy as distinct concerns, these distinctions have yet to be incorporated successfully in public discourse and policy deliberations about edtech ethical issues.

Policy responses to ethics of edtech

At the federal level, there has been some bipartisan congressional interest in issues related to edtech and student data privacy but no action. In 2015, eight education data privacy bills were introduced, four focused on regulation of schools and education agencies and three on regulation of third party companies (NASBE 2015). Among those seen as likely to gain support were Representatives Todd Rokita (R-IN) and Marcia Fudge's (D-OH) 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) and Senators Orrin Hatch (R-UT) and Edward Markey's (D-MA) amendment to the Elementary and Secondary Education Act, creating a Student Data Privacy Policy Committee with responsibility for studying and providing recommendations on privacy safeguards and parental rights. None of these moved beyond committee consideration but congressional interest in student privacy continues. On May 17, 2018 the US House Education and Workforce Committee held a hearing on "protecting privacy, promoting data security: exploring how schools and states keep data safe," at which witnesses agreed on the need to update FERPA and to address the complex issues presented by use of edtech at the K-12 level. However, there continues to be differences about how best to hold edtech vendors accountable and differences about the problems presented. For example, one commentator criticized the witness from the Future of Privacy Forum for praising the privacy policies of Class Dojo software, which the commentator characterized as "social emotional learning and behavioral modification software developed to inculcate, assess, and change students' personality traits in order to predict and steer children into careers chosen by corporations and governments, not the students" (Effrem 2018).

Absent congressional action, the policy focus instead has primarily been at the state level where issues have similarly been defined primarily as student privacy. Between 2013 and 2017, 503 bills addressing the privacy and security of education data were introduced in 49 states and 41 states have passed 94 new laws (DQC 2017). 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). The Data Quality Campaign identified two overlapping approaches in these

state bills: 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); and the *governance approach*, which established procedures (e.g., audits and inventories), roles and responsibilities to ensure appropriate student data practices. Most states have adopted the governance approach rather than the prohibitive approach. Bills that have restricted the use of data for student learning innovations, such as predictive analytics, have not been successful. States, like the federal government, seem reluctant to be overly prescriptive in ways that might hinder new technology applications with benefits for student learning or in ways where laws are written so specifically that technology advances outpace the laws (Roscorla 2016). Additionally, states are still sorting out the appropriate roles of state boards of education, school districts, and school boards (Regan and Khwaja 2017).

In addition to legislative deliberations, edtech companies and advocacy groups have adopted a degree of self-regulation, again defined along lines of protecting student privacy. The "K-12 School Service Provider Pledge to Safeguard Student Privacy" was initiated by the Future of Privacy Forum (FPF) and the Software and Information Industry Association (SIIA) in October of 2014 and as of August 2018 has been signed by 347 companies. Compliance with the Pledge opens signatories to the possibility of enforcement by the Federal Trade Commission as it is a public statement of the company's policy, similar to privacy notices. The Pledge provides the standard fair information practice principles of notice, choice, consent and transparency, as well as promises regarding security and prohibitions on use of information for advertising. The one commitment beyond the usual list is "Not build a personal profile of a student other than for supporting authorized educational/school purposes or as authorized by the parent/student" (FPF and SIIA 2018); however, the caveat of "authorized educational/school purposes" is not defined and may include incorporation of predictive analytics as has been raised in policy discussions regarding FERPA updates as well (Zeide 2017).

Policy discussions about the ethical issues raised by edtech applications are likely to continue and as we saw above are most often framed as privacy issues. At this point in the policy debate, we can identify a number of trends. The first is that the current focus is primarily on: the security of student data especially given reported data breaches and cyber attacks (Davis 2015; Doran 2017); deidentification of student data for research and analytical purposes; prohibitions on targeted advertising using student data; ownership of student information, with general agreement that ownership should remain with the local school district; and transparency regarding online practices, including inventories of online and edtech programs. 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, under both the Obama and Trump administrations, (including the 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 data to increase student success with a more flexible view on data privacy.

The second trend is that the policy discourse regarding these ethical issues is shaped primarily by the current legal framework and the standard fair information principles that have been incorporated in US privacy legislation in the past. 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. A recent Workshop Summary of the National Academy of Education states that "concerns about data privacy often focus on student information falling into the wrong hands or being used for nefarious purposes" and that "confidentiality and security flow from privacy" (2017, p. 4) and goes on to examine the key legislation protecting privacy. Such an approach fairly quickly locks the policy discussion into pre-existing categories, emphasizing fair information practices of notice, choice, consent and transparency thus quickly channeling identification of problems and solutions on a familiar path for policymakers and precluding 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. As Boninger et al. similarly conclude: "'personalized learning' software is slipping through the policy framework" (2017, p. 27). The next section of the paper discusses the potential of big data applications to discriminate through algorithmic biases, particularly personalized learning systems, 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" (Tucker 2015). 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 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 applied in some form throughout the K-12 level in most American schools – and has been controversial for more than a century (Loveless 2013; Burris and Garrity 2008, Chap. 2) .

The debate about tracking began in the late 1800s 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 the "Committee of Ten," appointed by the National Education Association (NEA), chaired by the 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 consider the variety of abilities, goals and financial means of the more diverse student population resulting from immigration. As a result, by the mid-1920s most 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 (Mirel 2006).

During the 1960s, the racial and class effects of student tracking received much attention and criticism—and these concerns persisted through the twentieth century with the consensus being that "tracking has minimal effects on learning outcomes and profound negative effects on equity outcomes" (Strauss 2013). 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" (Kohli 2014). Several factors account for the persistence of tracking 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" (Kohli 2014). As one parent noted, "You see kids entering the building through the same door…But the second door is racially stratified" (Kohli 2014).

The debate about student tracking has continued into the twenty-first 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" (2013).

Big data applications in education signal yet another fundamental change in the dynamics of sorting students. The actions of today's "digital student" 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 high school and beyond. Our review of the edtech 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, as well as attention spans). All this data 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 pursue.

Personalized learning systems represent the edtech application that raises the potential for tracking students along lines similar to the 1950s tracking of students in discriminatory ways. As Monica Bulger describes it:

> For many personalized learning systems, student data such as age, gender, grade level, and test performance are analyzed against idealized models of student performance, or students of the same background or class, or nationwide pools of grade and/or competency level. A profile is created for each student that typically categorizes her or him as part of a group that performs similarly or demonstrates shared interests or demographics. Then, data-driven content recommendations

> are sent either directly to the student or to the teacher for further intervention (2016, p. 2).

The possibility for socially unaccepted discrimination enters into such systems depending upon the student data used, the idealized model used, the profile used to categorize a student, and the recommendations offered. Although edtech personalized learning systems may eliminate human bias from educators in the school, it does introduce possible bias of those who design the systems and machine bias as the systems learn and evolve. For some time, scholars and some advocacy groups have recognized the possibility of bias in several contexts where predictive analytics and machine learning are being used, especially in policing, financial services, employment and, more recently, education. The policy responses to possible discrimination as a result of predictive analytics are still being explored. As Barocas and Selbst argue with respect to possible discrimination resulting from such systems in employment, existing anti-discrimination laws may not provide redress and may indeed appear to condone some practices (2016). And as we saw above, existing fair information practices laws also do not provide an effective solution to possible discrimination.

Discussion

As policymakers and scholars deliberate about the ethical dimensions of personalized learning and predictive analytics in the context of K-12 education, three factors appear critical to arriving at a genuine understanding of these ethical dimensions in order to fully inform policy choices.

First, the issues raised by personalized learning include all six of the ethical concerns discussed above and each concern needs to be addressed separately, rather than lumped under a "privacy" umbrella. Personalized learning programs amass a great deal of information not only about a student's responses to questions posed by the programs but also to the ways in which students interact with the program. Some notice, consent, and transparency about this should be provided to the schools and to parents. Standard FIPPs notices at this time do not to include this level of detail but inventories of personalized learning programs used at a specific school can be expanded to list such detail. Protecting students from reidentification in data sets or algorithmic searches resulting from or used in the creation of personalized learning programs is a different ethical issue and should be treated as such. Similarly, surveillance of students as they interact with these programs both in and out of the school environment raises ethical issues that are distinct from those addressed by FIPPS as knowledge of such surveillance may change the ways in which students respond to the programs or interact with one another or with a teacher and may normalize the expectation of surveillance in other areas of students' lives. The predictive analytics that are incorporated in many personalized learning programs may restrict the options available to students and thus limit the autonomy of students and of teachers who often do not understand or cannot easily explain why certain students are receiving different options than other students. The fact that differences in options emerge for different students entails discrimination among students and such differences may manifest or involve racial, ethnic, gender or other classifications that are not permitted by social norms or laws. Finally, questions about who owns the detailed information and analysis about student interactions with personalized learning programs is a separate ethical question that becomes further complicated as it involves the financial interests and incentives of edtech companies as well as the development of new edtech applications that may enhance student learning.

Second, more clarity about what is included under the rubric of "personalized learning" is needed. This has become something of a catch-all term but in the same way that "privacy" catches a variety of distinct interests and concerns, there are important distinctions among what are now termed "personalized learning" programs. Bulger developed a typology of technologically-enabled personalized learning systems beginning with customized learning interfaces; learning management platforms; data-driven learning platforms; adaptive learning platforms; and intelligent tutor platforms (2016, pp. 6–11). As one progresses along this continuum, computer-assisted programming appears to play a greater role in directing the learning process and the ethical issues need to be analyzed at each point to determine if and how they are qualitatively different than what occurs in a regular classroom setting with teachers assessing individual student progress .

Identifying differences in programs that are often characterized as "personalized learning" are important in order to determine the ethical issues posed by a particular program. Blended learning programs or hybrid programs may involve fewer ethical issues as they generally involve in-class instruction directed by a teacher supplemented by computer programs that the teacher has some control over, or clear understanding of, and in which the student exercises some control, especially over the path and pace of instruction (Horn and Staker 2011; Horn et al. 2013). Compromises of student and teacher autonomy are minimized in blended learning programs as both parties retain agency or control rather than turning that over to a computerized program. Tutoring programs that are designed to meet the needs of a student, are selected by a parent often upon advice of a teacher, and are produced by an edtech vendor that is not associated with programs used in the classroom, are also not likely to raise significant ethical issues if the data used by, collected by and analyzed by the program would not automatically be integrated with data collected by learning

programs used in the classroom. But four features appear critical to determining whether the level of ethical concern presented by personalized learning programs increases: first, as parent or teacher control or understanding of the learning software decreases, concern increases; second, as integration of the data collected by edtech vendors and classroom activities increases, ethical concern increases; third, as the edtech vendor collects data about more aspects of a student, concern increases; and fourth, as that data is used to refine software programs, ethical concern increases.

Finally, the *third* challenge to arriving at a genuine understanding of these ethical dimensions is that gathering data to inform ethical and policy decisions is enormously difficult. Much of this information, especially information about how algorithms are developed and how they evolve, is proprietary and held tightly by the edtech companies. Legislative efforts to regulate or restrict the use of predictive analytics have not been successful, meeting with opposition from the edtech industry. Parents do not have adequate information about software packages and tests to ask questions—and efforts to require inventories of edtech applications used at schools often result in more summary, and less useful, information. But the most significant problem with gathering data appears to be the tight relationships that are being established both between edtech companies and schools and between edtech companies and researchers; relationships which may raise possible conflicts of interest. For example, the Chan Zuckerberg Initiative's education division, which is separate from Facebook but somewhat related, has given millions of dollars to the Summit Learning Platform (marketed on its website as "free for teachers and schools" and that "helps students set and track goals, learn content at their own pace and complete deeper learning projects"), Summit Schools and AltSchool (founded by a former Google executive)—two of the major alternative schools using big-data approaches and personalized learning applications. In terms of research on personalized learning, the Bill and Melinda Gates Foundation has funded the three major RAND reports, has given more than $300 million to edtech initiatives, and has supported past coverage of personalized learning in *Education Week*, a major magazine for K-12 educators (Herold 2017). The actions of these venture philanthropists are serving to blur boundaries in public education between the public sphere and the private sphere in ways that are "shaping social perception about public education" and what is seen as necessary ways to reform public education (Baltodano 2017, p. 1520).

Conclusion

In order to fully and effectively address the ethical challenges posed by edtech applications, particularly applications falling into the personalized learning category, it is important not to oversimplify the discussions by grouping all concerns under the broad category of privacy. Additionally, ethical concerns should be fully vetted and explored in policy discussions. Being clear about the rationales for categorizing edtech applications is likewise warranted. The analysis above supports the position that policy based on fair information practice principles is an inadequate approach to address the range and nature of these concerns. Instead policy responses that open up the "black box" of what edtech applications actually do in terms of collecting information and how algorithms analyze that information are required. In this respect, our conclusions mirror the National Education Policy Center's policy recommendations that algorithms be available for examination by educators and researchers, that a disinterested third party review software using algorithms prior to adoption to ensure no bias or error, and that there be third-party reviews of the validity and utility of edtech prior to adoption (Boninger et al. 2017, p. 29). These recommendations are similar to those of the AI Now Institute in its 2017 report which noted the need for a "more intentional approach to ethics" and emphasized pre-release trials of artificial intelligence (AI) systems for 'high stakes' domains, including education, to ensure they do not include bias or error; standards for auditing of AI systems; and ethical codes accompanied by strong oversight and accountability (Campolo et al. 2017, pp. 33 and 1–2).

Adoption of edtech applications continues throughout K-12 education despite the lack of reliable evidence of their effectiveness in improving student learning. The need for more research to determine the impact of edtech applications on student achievement has been broadly recognized and such research is underway. In 2015 the OECD conducted a comparative analysis of student access to and use of information and communication technology (ICT) which concluded there was "no appreciable improvements in student achievement in reading, mathematics or science in the countries that had invested heavily in ICT for education" (OECD 2015, p. 3) The report emphasized that: "We need to *get this right* [emphasis added] in order to provide educators with learning environments that support twenty-first century pedagogies and provide children with the twenty-first century skills they need to succeed in tomorrow's world" (OECD 2015, p. 4). "Getting this right" does not just entail examining the effectiveness of the technology but also examining and fully addressing the ethical issues. At this point in the adoption of edtech applications, there is a danger that passage of laws framed as protecting student privacy and addressing only some of the ethical concerns will give the public a false sense that there are no other ethical considerations. The result could be that a generation of students is subtly tracked into learning paths that machine learning algorithms, rather than teachers or parents, decide are best for them and that

may discriminate among students in ways society and educators might not currently understand or support.

Acknowledgements The research for this paper was funded by the eQuality Partnership Grant from the Social Science and Humanities Research Council of Canada. For information on the grant, please see: http://www.equalityproject.ca/. An earlier version of this paper, "Big Data in the Education Arena: 21st Century Student Sorting and Tracking" co-authored by Priscilla M. Regan, Jolene Jesse and Elsa Talat Khwaja, was presented at the Surveillance Studies Network Conference in April 2016. The author acknowledges the assistance of Elsa Talat Khwaja in preparation of this paper. This material is based upon work supported while employed at the US National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Alarcon, A., Zeide, E., Rosenblat, A., Wikelius, K., boyd, d., Gangadharan, S. P., & Yu, C. (2014). Data & Civil Rights: Education primer, produced for Data & Civil Rights Conference. Accessed March 15, 2016, from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2542268.

Baltodano, M. (2017). The power brokers of neoliberalism: Philanthrocapitalists and public education. *Policy Futures in Education, 15*(2), 141–156.

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (Eds.), *Privacy, big data, and the public good: Frameworks for engagement* (p. 44). New York: Cambridge University Press.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review 104*, 671–732.

Bennett, C. J. (1992). *Regulating privacy: Data protection and public policy in Europe and the United States*. Ithaca: Cornell University Press.

Bienkowski, M. (2017). *Implications of privacy concerns for using student data for research: Panel summary*. Workshop on Big Data in Education. Accessed March 4, 2018, from https://naeducation.org/wp-content/uploads/2017/05/Bienkowski-FINAL.pdf.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational analytics and data mining*. Center for Technology and Learning, SRI International. Accessed March 1, 2018, from https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf.

Bogle, A. (2014). What the failure of InBloom means for the student-data industry. *Slate Future Tense Blog*. Accessed March 8, 2016, from http://www.slate.com/blogs/future_tense/2014/04/24/what_the_failure_of_inbloom_means_for_the_student_data_industry.html.

Boninger, F., Molnar, A., & Murray, K. (2017). *Asleep at the Switch: Schoolhouse commercialism, student privacy, and the failure of policymaking*. National Education Policy Center: Report on Schoolhouse Commercialization Trends. Accessed February 1, 2018, from http://nepc.colorado.edu/files/publications/RB%20Trends%202017_2.pdf.

Bulger, M. (2016). *Personalized learning: The conversations we're not having*, Working Paper 07.22.2016. Data and Society Research Institute. Accessed February 3, 2018, from https://datasociety.net/pubs/ecl/PersonalizedLearning_primer_2016.pdf.

Bulger, M., McCormick, P., & Pitcan, M. (2017). *The legacy of inBloom*, Working Paper 02.02.2017. Data and Society Research Institute. Accessed February 3, 2018, from https://datasociety.net/pubs/ecl/InBloom_feb_2017.pdf.

Burris, C. C., & Garrity, D. T. (2008). *Detracking for excellence and equity,"* see especially Chap. 2 "What Tracking is and How to Start Dismantling It." Accessed March 15, 2016, from http://www.ascd.org/publications/books/108013/chapters/What-Tracking-Is-and-How-to-Start-Dismantling-It.aspx.

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI Now 2017 Report*. Accessed March 3, 2018, from https://ainowinstitute.org/AI_Now_2017_Report.pdf.

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review, 89*, 101–133.

Cohen, J. E. (2000). Examined lives: Informational privacy and the subject as object. *Stanford Law Review, 52*, 1373–1438.

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review, 55*, 93–128.

Data Quality Campaign. (2015). *State student data privacy legislation: What happened in 2015, and what is next?* Accessed February 15, 2016, from https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2016/03/DQC-Student-Data-Laws-2015-Sept23.pdf.

Data Quality Campaign. (2017). *Education data legislation review: 2017 state activity*. Accessed January 28, 2018, from https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2017/09/DQC-Legislative-summary-0926017.pdf.

Davis, M. R. (2015). Lessons learned from security breaches. *Education Week, 35*(9), S6–S7.

Doran, L. (2017). Ransomware attacks force school districts to shore up–or pay up. *Education Week, 36*(17), 1–10.

Effrem, K. R. (2018). 6 Key takeaways from Congress' hearing on protecting student data. *The National Pulse*. Accessed August 15, 2018, from https://thenationalpulse.com/commentary/6-key-takeaways-congress-hearing-protecting-student-data/.

English, W. (2016). Two cheers for nudging. *Georgetown Journal of Law and Public Policy, 14*, 829–840.

Flaherty, D. (1989). *Protecting privacy in surveillance societies: The Federal Republic of Germany, Sweden, France, Canada, and the United States*. Chapel Hill: University of North Carolina Press.

Future of Privacy Forum and Software & Information Industry Association. (2018). *K-12 school service provider pledge to safeguard student privacy*. Accessed August 20, 2018, from https://studentprivacypledge.org/privacy-pledge/.

Gellman, R. M. (1993). Fragmented, incomplete, and discontinuous: The failure of federal privacy regulatory proposals and institutions. *Software Law Journal, 6*, 199.

Hartzog, W., & Selinger, E. (2013a). Big data in small hands. *Stanford Law Review Online, 66*, 81–88.

Hartzog, W., & Selinger, E. (2013b). Obscurity: A better way to think about your data than privacy, *Atlantic*. Accessed January 10, 2016, from http://www.theatlantic.com/technology/archive/2013/01/obscurity-a-better-way-to-think-about-your-data-than-privacy/267283/.

Herold, B. (2014). InBloom to shut down amid growing data-privacy concerns, *Education Week*. Accessed March 8, 2016, from http://blogs.edweek.org/edweek/DigitalEducation/2014/04/inbloom_to_shut_down_amid_growing_data_privacy_concerns.html.

Herold, B. (2017). Personalized learning: Modest gains, big challenges, RAND study finds, *Education Week*. https://blogs.edweek.org/edweek/DigitalEducation/2017/07/personalized_learning_research_implementation_RAND.html. Accessed 10 February 2018.

Horn, M. B., & Staker, H. (2011). *The rise of K-12 blended learning*. Innosight Institute. Accessed January 20, 2018, from https://www.christenseninstitute.org/publications/the-rise-of-k-12-blended-learning/.

Horn, M. B., Staker, H., & Christensen, C. (2013). *Is K-12 blended learning disruptive? An introduction to the theory of hybrids*. Accessed January 20, 2018, from https://www.christenseninstitute.org/publications/hybrids/.

Kelly, J. T. (2016). Non-paternalist nudges. *Georgetown Journal of Law and Public Policy, 14*, 807–816.

Kerr, I., & Earle, J. (2013). Prediction, preemption, presumption: How big data threatens big picture privacy. *Stanford Law Review Online, 66*, 65–72.

Kohli, S. (2014). Modern-day segregation in public schools, *The Atlantic*. Accessed March 1, 2016, from http://www.theatlantic.com/education/archive/2014/11/modern-day-segregation-in-public-schools/382846/.

Lerman, J. (2013). Big data and its exclusions. *Stanford Law Review Online, 66*, 55–63.

Lichtenberg, J. (2016). For your own good: Informing, nudging, coercing. *Georgetown Journal of Law and Public Policy, 14*, 663–682.

Loveless, T. (2013). The resurgence of ability grouping and persistence of tracking, (Part II of the 2013 Brown Center Report on American Education), *Brookings Report*. Accessed March 3, 2016, from http://www.brookings.edu/research/reports/2013/03/18-tracking-ability-grouping-loveless.

Mirel, J. (2006). The traditional high school: Historical debates over its nature and function, *Education Next* (Winter) Vol 6, No. 1, pp. 14–21. http://eric.ed.gov/?id=EJ763310. Accessed 1 March 2016.

National Academy of Education Workshop Summary. (2017). *Big data in education: Balancing the benefits of educational research and student privacy*. Accessed February 23, 2018, from http://naeducation.org/wp-content/uploads/2017/05/NAEd_BD_Booklet_FINAL_051717_3.pdf.

National Association of State Boards of Education. (2015). *Comparison of 2015 Federal Education Data Privacy Bills*. Accessed August 15, 2018, from http://www.nasbe.org/wp-content/uploads/2015-Federal-Education-Data-Privacy-Bills-Comparison-2015.07.22-Public.pdf.

New, J. (2016). *Building a data-driven education system in the United States*. Center for Data Innovation. Accessed August 15, 2018, from http://www2.datainnovation.org/2016-data-driven-education.pdf.

Ohm, P. (2014). General principles for data use and analysis. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (Eds.), *Privacy, big data, and the public good: Frameworks for engagement* (pp. 96–111). New York: Cambridge University Press.

Organization for Economic Cooperation and Development (OECD). (2015). *Students, computers and learning: Making the connection*. Programme for International Student Assessment, OECD Publishing. Accessed August 16, 2018, from https://read.oecd-ilibrary.org/education/students-computers-and-learning_9789264239555-en#page1.

Pane, J. F., Steiner, E. D., Baird, M. D., Hamilton, L. S., & Pane, J. D. (2017). *Informing progress: Insights on personalized learning implementation and effects*. RAND Report. Accessed February 13, 2018, from https://www.rand.org/pubs/research_reports/RR2042.html.

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. New York: Penguin Books.

President's Council of Advisors on Science and Technology. (2014). *Big data and privacy: A technological perspective*. Accessed December 10, 2015, from http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

Regan, P. M. (1995). *Legislating privacy: Technology, social values, and public policy*. Chapel Hill: University of North Carolina Press.

Regan, P. M. (2017). Big data and privacy. In J. Bachner, K. W. Hill, & B. Ginsberg (Eds.), *Analytics, policy and governance*. New Haven: Yale University Press.

Regan, P. M., & Bailey, J. (2018). Big data, privacy and education applications. In *Presented at Surveillance Studies Network Conference, June 2018*. Denmark: Aarhus University.

Regan, P. M., & Khwaja, E. T. (2017). Ethical implementation of big data in education: Policy and practices in the US and Canada. In *Presented at the Law and Society Association Annual Conference, June 2017*, Mexico City.

Roberts-Mahoney, H., Means, A. J., & Garrison, M. J. (2016). Netflixing human capital development: Personalized learning technology and the corporatization of K-12 education, *Journal of Education Policy* 1–16.

Roscorla, T. (2016). 3 Student Data Privacy Bills that Congress Could Act On, *Government Technology: Center for Digital Education*. Accessed August 15, 2018, from http://www.govtech.com/education/k-12/3-Student-Data-Privacy-Bills-That-Congress-Could-Act-On.html.

Rubinstein, I. S. (2013). Big Data: The end of privacy or a new beginning? *International Data Privacy Law, 3*(2), 74–87.

Schwartz, P. M. (2000). Internet privacy and the state. *Connecticut Law Review, 32*, 815.

Schwartz, P. M. (2004). Property, privacy, and personal data. *Harvard Law Review, 117*(7), 2055–2128.

Singer, N. (2013). Deciding Who Sees Students' Data, *The New York Times*. Accessed February 16, 2016, from http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html.

Singer, N. (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 8 March 2016.

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

Strauss, V. (2013). The bottom line on student tracking, *The Washington Post*. Accessed March 8, 2016, from https://www.washingtonpost.com/news/answer-sheet/wp/2013/06/10/the-bottom-line-on-student-tracking/.

Sunstein, C. R. (2015). The ethics of nudging. *Yale Journal on Regulation, 32*(2), 413–450.

Sweeney, L. (2000). *Uniqueness of simple demographics in the US population* (Laboratory for International Data Privacy, Working Paper LIDAP-WP4). Accessed December 10, 2015, from http://dataprivacylab.org/projects/identifiability/index.html.

Sweeney, L., Abu, A., & Winn, J. (2013). Identifying participants in the personal genome project by name, *Harvard University Data Privacy Lab*, White Paper 1021-1 (April 24). Accessed December 10, 2015, from http://dataprivacylab.org/projects/pgp/1021-1.pdf.

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property, 11*(5), 239–273,253.

Tucker, M. (2015). Student tracking vs academic pathways: Different… or the same? *Education Week* (October 15). Accessed March 8, 2016, from http://blogs.edweek.org/edweek/top_performers/2015/10/tracking_vs_pathways_differentor_the_same.html.

West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards, *Governance Studies at Brookings* (September). Accessed March 5, 2016, from http://www.brookings.edu/~/media/research/files/papers/2012/9/04-education-technology-west/04-education-technology-west.pdf.

Westin, A. (1967). *Privacy and freedom*. New York: Atheneum.

Zeide, E. (2016). Student privacy principles for the age of big data: Moving beyond FERPA and FIPPS. *Drexel Law Review, 8*, 101–160.

Zeide, E. (2017). The limits of education purpose limitations. *University of Miami Law Review, 41*, 494–527.