Student Privacy and Educational Data Mining: Perspectives from Industry

Jennifer Sabourin Lucy Kosturko Clare FitzGerald Scott McQuiggan
Curriculum Pathways
SAS Institute Inc.
100 SAS Campus Drive
Cary, NC, USA
1.919.677.8000
{Jennifer.Sabourin, Lucy.Kosturko, Clare.FitzGerald, Scott.McQuiggan}@sas.com

ABSTRACT
While the field of educational data mining (EDM) has generated many innovations for improving educational software and student learning, the mining of student data has recently come under a great deal of scrutiny. Many stakeholder groups, including public officials, media outlets, and parents, have voiced concern over the privacy of student data and their efforts have garnered national attention. The momentum behind and scrutiny of student privacy has made it increasingly difficult for EDM applications to transition from academia to industry. Based on experience as academic researchers transitioning into industry, we present three primary areas of concern related to student privacy in practice: policy, corporate social responsibility, and public opinion. Our discussion will describe the key challenges faced within these categories, strategies for overcoming them, and ways in which the academic EDM community can support the adoption of innovative technologies in large-scale production.

Keywords
Student privacy, student data, policy

1. INTRODUCTION
Educational data mining (EDM) is chiefly defined by the application of sophisticated data mining techniques to solving problems in education [1]. A powerful tool, EDM has been successfully incorporated into applications that optimize student learning in both research and commercial products. EDM's proven effectiveness has led many—from the U.S. government to individual teachers—to recognize the ability of student data in guiding education and to support the development and use of these technologies in schools. Consequently, applications utilizing EDM technologies have become more prevalent in school systems [2], [3].

However, the increase in EDM usage has raised public awareness of how much data is being collected about students. The applications and companies that collect and use student data are coming under scrutiny, as parents, advocates, and public officials grow concerned over student privacy. A recent cascade of events has focused attention on privacy concerns [4]. For example, there has been a rise in high-profile attacks on consumer data from online retailers and financial institutions. Large, well-trusted institutions have been targeted for using student data in undesirable ways [5]. Promising companies driven by student data have been brought down by public opinion with no evidence of wrong-doing. Calls for stricter policy from privacy advocates have led to more than 100 bills being introduced in U.S. state legislatures to address issues of student privacy in 2014 [4]. In response, the White House has announced plans for federal legislation modeled after state policies [6].

Negative media attention and increased legislation threaten to stifle EDM, particularly in commercial settings. Public opinion may make organizations wary to invest in and use EDM techniques while legislation could make it more difficult to collect and use student data in effective ways. We believe it is an incredibly important time for the EDM community to be aware of the challenges being faced in industry. The rise of concern over student privacy has strong implications for how new EDM approaches can be integrated into wide-reaching applications as well as the amount of funding available to public and private entities wishing to innovate in this space.

These issues are receiving rapidly increasing attention and driving action at the national level. It is critical that the discussions around these issues include experts from the EDM community. This paper discusses the issues and implications faced by commercial applications of educational data mining because of recent focus on student privacy. In this paper, we discuss the role of policy, corporate social responsibility, and public opinion in framing the work of and challenges to industry. We discuss strategies for overcoming these challenges and present opportunities for the EDM community to address rising concerns.

2. EDM AND INDUSTRY
The profile of the EDM community has risen in the past decade—in research, commercial products, public attention—bolstered by three related shifts. First, educational technology has been more widely adopted. School systems are investing in laptops, mobile devices and other technologies in favor of static textbooks. These technologies offer opportunities for data collection that did not exist before. Student records are also increasingly digitized including test scores, attendance records, and bus schedules. These digitized records have generated a wealth of longitudinal data that was previously difficult and expensive to collect [7].

Second, there has been a dramatic rise in computational power and storage capacities. This storage allows for the collection and housing of large amounts of data, even data that is not presently known to be useful. The increased computational power has generated sophisticated algorithms that can mine large corpora of data to identify connections that would previously be impossible [8] and has even created the possibility for robust decision engines to operate in real time learning systems.

Finally, public officials and industry experts are starting to recognize the power of educational data mining [9]. Government funding opportunities for data-driven education solutions are on the

rise, and reports estimate that educational data mining has the potential to provide meaningful economic impact worldwide [10].

There are many areas of EDM research, each with unique applications to industry. At the individual level, data on student behavior, from mouse clicks to eye tracking, provide insight on how students interact with educational technology. For example, EDM has produced models of help abuse [11], attention to hints [12], and conversational dynamics in online forums [13]. These insights and techniques can help commercial educational technology providers design better applications that support positive interactions with students while being user-friendly.

Another key area of research at the individual level is assessment. EDM applications have been used to identify student mastery as well as knowledge gaps. Frequently, these models are based on student performance on relevant tasks but can go beyond measuring what students did correctly and incorrectly by modeling underlying knowledge [14]. Some assessments are cleverly hidden, called "stealth assessment," in games or other non-threatening applications [15]. These systems develop robust models of student knowledge while avoiding the negative effects associated with test performance; in fact, students may not even knowing they are being tested. These techniques have important implications for educational technologies, ranging from the design of new systems that can revolutionize the way assessment is done in formal learning environments, to technologies that can identify gaps in student knowledge and recommend resources to help fill them.

EDM technologies have also driven personalized learning beyond tailoring instruction to what students know, but also to how they learn based on needs and preferences. Systems can identify commonly used strategies by students and select which are most effective, for particular individuals, under specific circumstances [16]. EDM techniques have also supported technologies that guide students towards learning how to regulate their own learning, by helping them to recognize and overcome weaknesses in their current approaches [17]. These techniques are critical in creating applications that use the most effective techniques and support personalized learning.

Finally, EDM research has examined mining data at higher levels, including schools and districts, for a variety of purposes such as exploring college readiness [18], identifying the best teachers [19], or driving district spending [7]. Commercial products are commonly used to house this level of data and communicate findings to necessary stakeholders. Data mining on this organizational or even regional level has allowed for the development of early warning systems to predict student drop-out before it happens as well as identify holes in district-level education [7].

In essence, "educational data mining and learning analytics have the potential to make visible data that have heretofore gone unseen, unnoticed, and, therefore, unactionable" [9]. The approaches outlined in this section offer significant promise in helping to improve education delivery and outcomes, but their success is contingent on the collection, storage, and use of large amounts of quality student data. Companies who wish to collect and use student data must operate under increased public and governmental scrutiny, which can, and has, created barriers to the use of EDM in industry.

3. STUDENT PRIVACY

Privacy is chiefly a question of access. Unlike anonymity or confidentiality, peoples' interest in privacy is about controlling the access of others to themselves [20]. How to safeguard a child's privacy is a particularly complex question because of their vulnerability. Children are incapable of "protecting their own interests through negotiation for informed consent" because they are likely to misunderstand risks or be coerced into participating [20].

This need to protect has led to the formation of student privacy advocacy groups and driven the adoption of legislation. The restrictions required to comply with this legislation and maintain good public opinion have a significant impact on the adoption of data-based solutions in education.

3.1 Policy

In the U.S., we have established privacy protections for children by asking for consent from parents or guardians and implementing policies which hold organizations, both public and private, accountable for obtaining consent when collecting, storing or disclosing data, and ensuring proper usage. There are two federal acts that address children's privacy directly: the Federal Education Rights and Privacy Act (FERPA), and the Children's Online Privacy Protection Act (COPPA).

3.1.1 Federal Education Rights and Privacy Act

Before the enactment of the Federal Education Rights and Privacy Act (FERPA) in 1974, parents and students had little access to education records. Meanwhile, that same information was widely available to outside authorities without requiring the consent of parents or students [21]. FERPA applies to any school receiving federal funds and levies financial penalties for not following it. While complying with FERPA is a local responsibility [22], the way it defines education records and regulates third party access to them matters to private companies.

According to FERPA, education records contain information on student background, academic performance, grades, standardized test results, psychological evaluations, disability reports, and anecdotal remarks from teachers or school authorities regarding academic performance or student behavior (FERPA, 1974, 20 U.S.C § 1232g(a)(1)(D)(3). Generally, schools looking to disclose information contained in these records must have written permission from a parent or eligible student, an individual who is 18 or attending post-secondary school. Education record information is only shared with a third party on the assurance that that third party will not allow further outside access to requested information without additional written parental consent (FERPA, 1974, 20 U.S.C § 1232g (b)(4)(B)). Some activities, however, do not require written consent. Under FERPA, third parties, including private companies, may use information within education records for official or contracted evaluation, audit, and compliance activities without parental or student consent but are barred from using that data for marketing [23].

FERPA is not without controversy. Some have argued that schools improperly apply FERPA in order to protect information that does not fall under its definition of an education record and that such denials of disclosure are in violation of state open record laws [24]. Others voice concern over contracted service providers' use of data not covered by FERPA citing that the content of emails housed in cloud services, data from identification cards, or data collected by schools to outsource a service could, depending on the contract, be used or sold for marketing purposes [23].

3.1.2 Children's Online Privacy Protection Act

While FERPA affects private interests, the Children's Online Privacy Protection Act (COPPA) speaks more directly to

operations, particularly to online service providers that have direct or actual knowledge of users under 13 and collect information online. Made effective in 2000, COPPA "requires web hosts and content providers to seek parental consent to store data about children under age 13" [25]. To be fully compliant, parents must be given the opportunity to review terms of service and privacy policies of each commercial website where their child's information may be stored. Parental consent is required before any information can be collected, and parents can retract this permission and request all data be deleted at any time. Technology providers are required to disclose what data is being collected about children and what it is being used for. They are also expected to provide reasonable measures of security and discard of data once it is no longer needed. [25], [26]. Overall, COPPA seeks to encourage responsible business practice and reduce "imprudent disclosures of personal information by children" [27].

COPPA, too, has fallen under criticism. It is difficult to enforce and there many ways in which companies can comply with the "letter of the law" without truly protecting student privacy. COPPA has also been criticized for not reflecting the changes in online technologies accessed by children. In an effort to stay current with technological advancement, COPPA underwent revisions in 2013 to "address changes in the way children use and access the Internet, including the increased use of mobile devices and social networking" [28] by widening the definition of what constitutes children's personal information to include cookies, geolocation, photos, videos, and audio recordings [28]. These updates bolstering safeguards for student data appear further scaffolded by actions from the White House.

3.1.3 Student Digital Privacy

Driven by concerns over the efficacy of national policies, state legislators have seen the introduction of a large number of policies aimed at protecting student data [4], [29]. New national legislation may also be on the horizon for protecting student privacy [30]. The proposed Student Digital Privacy Act, modeled after a California statute, prohibits companies from selling student data to third parties except for educational purposes [6]. While it is unclear when, or if, this legislation will be enacted, it has already drawn criticism. Parents and privacy advocates fear it is too lenient while industry experts warn that increased legislation may limit development of important educational solutions [31].

These industry experts point to the voluntary Student Privacy Pledge (http://studentprivacypledge.org/) as a means to achieve better management of student data without federal legislation [32]. At the time of writing, 108 companies have chosen to sign the pledge, vowing that they will not sell student data or use data for targeted advertisement, and will maintain transparency about how data is being collected and used. This pledge is an indication that commercial education technology providers are taking steps towards the corporate social responsibility that will garner respect among users and privacy advocates.

3.1.4 Student Privacy: International Perspectives

The United States has relied on a piecemeal approach to regulating privacy where legislation is sector driven and may be enacted at state and/or federal levels [33]. Conversely, the European Union enacted a comprehensive set of regulations in the Data Protection Directive under which student privacy issues are largely subsumed. This set of regulations requires unambiguous consent of individuals before collecting or processing personal data as well as a prohibition on collecting sensitive information with few exceptions [34].

Canadian national privacy legislation is stipulated in the Personal Information Protection and Electronic Documents Act which, like COPPA, is focused on how commercial entities use personal information, as well as the Privacy Act which limits the collection, use, and disclosure of personal information by federal government entities. Meanwhile, similar to United States, Canadian provinces follow their own patchwork of student specific legislation. Ontario, for instance, follows the Education Act, the Municipal Freedom of Information and Protection of Privacy Act as well as the Personal Health Information Protection Act. The Canadian system is less comprehensive than the EU, but is perhaps more effective in safeguarding student interests than the US due to an "all-encompassing and prescriptive nature" [34].

3.2 Corporate Social Responsibility

Corporate social responsibility refers to companies taking an active part ensuring they have a positive impact on social welfare. In the case of privacy, this means working to truly protect student data and collect and use it responsibly. Design weaknesses and enforcement shortcomings in student privacy legislation can often allow companies to appear more responsible than they are. Organizations can legally comply, a potentially cumbersome process on its own, but do little to actually ensure best practices are being followed and student interests are protected.

This is a significant issue in markets of educational technologies designed for children under the age of 13, the population protected by COPPA. True compliance with the intents behind COPPA can be "both overwhelming and prohibitive" [35] which privacy scholar, Danah Boyd, believes has led to an apprehension to target users under thirteen. Avoiding the issue is often seen as "easier and more cost effective than attempting to tackle COPPA compliance." [35]

Currently there are many websites, online services, and mobile apps that are widely used in classroom settings including those classrooms with younger students. For example, Google Apps for Education reportedly serves an estimated 40 million students, teachers, and administrators. Similarly, over 47 million teachers have accounts with Edmodo, the "world's largest K-12 social learning community". Education technology is estimated to be an 8 billion dollar industry [30] and technology providers are often trying to find their niche while maintaining competitive advantage. Issues arise when creating a product that will be useful to education, ensuring that student data is collected and managed responsibly, and managing profit and competition are at conflict with one another. This balance of constraints is one of the strongest challenges faced by companies seeking to gather and use educational data responsibly.

3.2.1 Supporting Shared-Device Settings

Classroom constraints make the educational market particularly unique. While 1:1 schools (1 device per student) and Bring Your Own Device (BYOD) integrations are on the rise, many schools reflect a shared-device model (e.g., classroom sets, device carts). In order to achieve personalized learning in this setting, individual accounts are often necessary. Yet individual accounts raise several issues.

The first is that secure account authentication can be troublesome. Expecting students, especially younger students, to remember their login credentials is unreasonable in many cases. Keeping up with login information is particularly challenging when classrooms attempt to take advantage of multiple systems each requiring their own unique username and password. In fact, a report by the National School Board Association notes "password reuse due to

lax controls (i.e., password written on a sticky note)" as a particular concern for using online educational services [36]. Some systems utilize password pictures or avatars for younger populations, which could be a viable option depending on the type of data; however, when sensitive data such as images, video, and performance evaluations are often protected behind account logins, it is important to enable users to securely protect their data.

Furthermore, for those companies without any interest in storing student data on servers, shared-device settings can unintentionally force this responsibility. In a 1:1 environment, user data can simply be stored on students' devices as there is little concern over other individuals gaining access to the data; thus, eliminating the need to device solutions for complying to privacy legislation and avoiding security breaches. Appealing to shared-device environments, on the other hand, necessitates such measures including cloud storage, a solution known to concern parents [37]. Moreover, when schools rely on online educational resources and mobile apps that utilize cloud storage, they often relinquish control of that student data, which is particularly alarming given the fact that FERPA "generally requires districts to have direct control of student information when disclosed to third-party service providers" [23]. A recent report by Fordham Law School on the issue of student privacy and cloud computing found "school district cloud service agreements generally do not provide for data security and even allow vendors to retain student information in perpetuity with alarming frequency" [23]. The report goes on to point out that "fewer than 25% of the agreements [pulled from a national sample and reviewed by the committee] specify the purpose for disclosures of student information, fewer than 7% of the contracts restrict the sale or marketing of student information by vendors, and many agreements allow vendors to change the terms without notice." In sum, supporting ubiquitous student access through cloud computing necessitates a great deal of legal accommodations.

3.2.2 Consent

The process for simply creating an account can be cumbersome and time-consuming for two primary reasons: 1) companies cannot collect personal information from students under thirteen without parental consent, and 2) students under 18 cannot legally agree to the Terms of Service agreement accompanying many registration processes. In some cases, schools obtain a blanket agreement from parents at the beginning of the year allowing instructors to create accounts for students. Although, if teachers do not have legal consent from parents to create accounts on their students' behalf, having to wait for parental approval can easily derail an entire lesson quickly making the resource obsolete to the instructor.

Unfortunately, many companies find "restricting" users, even audiences for which the product is intended, streamlines the registration process by avoiding parental consent. Susan Fox of the Walt Disney Company articulates this concern by stating, "Operators are keenly aware that consumers will quickly move on if websites are slow to load, functionality is delayed, or registration-type processes stand between users and their content." [38] Furthermore, because virtual age verification is difficult and easily bypassed, compliance can still be met by adding statements such as "we do not knowingly collect data" from persons under thirteen in privacy policies. As a result, sidestepping the intentions of COPPA makes it difficult for other companies to remain competitive and "discourage[s] startups from innovating for the under-thirteen market" [38].

3.2.3 Disclosure

Parental consent and disclosure are two of the major tenants of COPPA compliance. Responsible adherence suggests that companies are forthcoming with information and present details clearly to parents when asking consent. However, this can be troublesome and may serve to harm parental opinions of an application rather than help. For example, there is concern that anything requiring parental permission (e.g., PG-13 or R-rated movies) is somehow objectionable. This misconception stems from the fact that "parents and youth believe that age requirements are designed to protect their safety, rather than their privacy." [39] As a result, companies attempting to be compliant may be inadvertently penalized because of public opinion.

Privacy policies are another form of disclosure that may be open to misinterpretation. Regulated by the FTC, privacy policies require companies to be upfront about the collection and use of user data. There is, however, much debate about their effectiveness. In a recent survey, over half of interviewed online Americans agreed with the statement, "When a company posts a privacy policy, it ensures that the company keeps confidential all the information it collects on users" and even fewer users read—or, in the case of these younger populations, can read and comprehend—them [40]. Others have proposed alternative solutions that more clearly convey the purposes of data collection [41] yet truly articulating the intricacies of EDM and personalized learning environments will take proofs of concept and time.

3.3 Public Opinion

One of the largest drivers behind the focus on privacy of student data is the vocal concern of parents and stakeholders in the media. The issue has been gaining a great deal of attention and has already had serious impacts on the landscape of educational technology providers.

Perhaps one of the best examples of the power of backlash from parents and media is the demise of a well-funded nonprofit company based entirely on the promise of educational data mining [5]. Though it was widely supported by districts, industry experts, and funding agencies, its efforts were undermined by parental protests and media frenzy. The company did not respond to rising concerns and failed to staunch fears over data misuse and protection. Though there was no evidence of any wrong-doing on the part of the company, parents and privacy advocates protested that the risk was too great. As the protest grew larger and more vocal districts began withdrawing participation in early 2014.

While anecdotal, this example demonstrates the need for industries relying on student data to get ahead of the rising panic by demonstrating value (i.e. driving innovation and/or supporting student learning). While EDM has its proponents [2], [9], their beliefs do not propagate to the general public. Parents and privacy advocates do not believe the benefits to be gained by educational technologies driven by student data outweigh the risks. The top concerns for these individuals are varied, as are their levels of awareness with various issues. Commonly discussed areas of concern with regards to student data include marketing, security, decision-making, and the "unknown".

3.3.1 Marketing

A primary purpose behind existing and proposed legislation is to limit the use of children's data to drive targeted advertisements [42]. It is, therefore, unsurprising that this is one of the top concerns of parents and school officials. However, much of this legislation and parental concern stems from children's interactions with non-educational sites and technologies. In this case, it makes sense to

limit targeted advertising of toys, food items, and other commercial goods, especially when considering findings that children are mostly unable to distinguish advertisement from regular content [43].

However, it is not clear that this protection is warranted in educational contexts. Much of the "advertisement" promoted by the EDM community centers around identifying gaps in a student's understanding and surfacing the most effective and engaging ways to fill those gaps. These advertisements have strong potential to benefit students, but some parents and other privacy advocates are only able to see that their children are being exploited for profit [2][3].

3.3.2 Decision-Making

Several EDM technologies provide a promise to support data-driven decisions about how best to help students learn. This is seen regularly in tools that select problem sets, feedback, or lesson plans based on students' prior interactions [44]. Data may also be presented to educators or administrators making decisions about whether a student needs additional attention or if they are college-ready [18]. These types of decisions start drawing parental concern. While parents understand (though they may not agree with) data from high stakes examinations being used to drive decisions about their children's education, data from private learning technologies is more unclear. Parents fear that undisclosed "stealth assessments" could negatively impact their children's future – from academics through the work force [42].

3.3.3 Security.

In addition to concerns over what companies may do with the data they collect, many parents are also fearful over what may happen if that data enters the wrong hands. The news is rife with incidents of data breeches with individual financial and other personal data being accessed by malicious parties. Parents concerns over student data security is certainly valid, though experts think it unlikely that this type of data would draw attack as it is less obviously lucrative when compared with financial and other personal records [2].

Existing legislation does put restrictions on the collection and storage of personally identifiable information (PII) of minors and responsible companies do strive to ensure anonymization of data. However, the rapid increase in the quantity of data collected and the sophistication of data mining procedures increase the likelihood that data that does not seem like PII on the surface could be combined to identify individuals [8].

3.3.4 The "Unknown".

Finally, many fears from parents and the media cannot be vocalized. There is something unsettling about the quantity of data being collected, stored, and mined about children, even if there is no real threat to safety or happiness. Much of this fear stems from the lack of transparency that surrounds the issues. Companies want to keep practices secret to avoid giving competitors an advantage. Privacy policies are often vague and uninformative to reduce the risk of drawing criticism or lawsuits. This is especially a concern as media tensions and attacks rise. Parents know that large quantities of data are being collected about their children, and it is unclear why it is being collected, how it is being used, and what it could be used for in the future. Rising distrust between parents, stakeholders and technology providers shuts down constructive conversation and only serves to exacerbate the issue.

4. ROLE OF THE EDM COMMUNITY

The barriers to industry applications of educational data mining techniques stem from several sources. Existing and proposed policy put restrictions on how data can be collected, stored and used. Companies can technically comply with legislation without much impact on their product or processes. However, strictly adhering to policies and offering real privacy protection often makes accessing and using educational tools more difficult, giving less socially responsible companies a competitive advantage. Public opinion can lead to the destruction of companies with no unethical practices and can drive money away from investment in data-based educational technologies. The EDM community has an important role to play in keeping these challenges in check and allowing innovation to thrive (Table 1).

4.1 Transparency

A lack of clarity, rampant misunderstanding, and a high degree of uncertainty fuel sentiment against the collection and use of student data. The main concerns of many parents and privacy advocates are largely not reflective of actual practice.

Consequently, the EDM community is unique positioned to advance public understanding for what student data is really being used. EDM professionals can better describe how data is being used, what innovations it supports, explain the focus of current research, and portray likely research foci of the field. Parental concerns may be allayed knowing that people are not actively contributing to the outcomes they most fear.

The community can also disseminate details about the effectiveness of these approaches beyond the research community. Showing the strengths of these techniques may help concerned individuals see the benefits that individual children and the education system as a whole stand to gain.

As new approaches are developed, consider creating public-facing talking points that can be used to communicate with concerned parties. These points should describe what data is being used and

Table 1. The role of the EDM community on the issue of student privacy.

Point of Concern	Proposed Solution	Action Item
Policy	• Policy Activism	• Remain abreast of proposed or approved policy changes. • Actively voice expert opinions to policy makers.
Corporate Social Responsibility	• Awareness of classroom constraints	• Develop algorithms that minimize the amount of data needed to produce effective results where possible. • Avoid requirements for individual accounts when possible.
Public Opinion	• Understanding public opinion • Transparency	• Actively work to correct misconceptions about student data and privacy concerns. • Set research agendas aimed at better understanding public understanding of privacy issues.

how it can benefit students. They should be written in a way that is clear and easy for non-experts to understand.

4.2 Research Agendas

The EDM community can also drive research towards areas that may help compliance with legislation and improve public opinion. Algorithms that minimize the amount of data needed to produce effective results would be beneficial to companies wishing to keep privacy concerns at bay. Researchers should consider the tradeoffs when developing new "big data" approaches. More data may lead to more effective techniques but it also may represent an increased violation of privacy. Finding a balance can support widespread dissemination in commercial technologies

It is important that researchers understand the classroom constraints of commercial educational technologies, especially when it comes to privacy. For a variety of reasons it is often less feasible to guarantee that data comes from a specific individual. Approaches that are robust enough to take this into account will allow educational technologies to be successful in more environments.

An additional area of research that could benefit from the involvement of the EDM community is research on the public understanding of privacy issues. The EDM community could be involved in cross-disciplinary research to ensure that communication surrounding EDM techniques is accurate and clear, and organizational privacy policies are widely understood.

4.3 Policy Activism

Finally, we encourage members of the EDM community to become active as policy debates grow. It is important to stay up to date on proposed policy changes and to consider how these changes may impact research agendas and the commercial applicability of those findings. Policy changes may increase constraints in commercial applications that could drive shifts in funding made available to EDM research. The policy changes affect both communities.

The discussion also needs more contributions from EDM experts. Consider voicing concerns to local officials and provide guidance on how policy should be directed. Too much of the current dialogue is based on a fear and misunderstanding. These voices are currently overpowering the experts who support the use of data in education.

5. CONCLUSION

Educational data mining offers significant promise in improving student learning and education systems as a whole. However, these systems are often driven by the collection of large amounts of student data, which is a growing concern to many. Shifts in public opinion and policy have led to barriers to the adoption of EDM technologies in commercial applications and threaten to stifle future innovation. Several fundamental issues are driving this trend.

The first is the role of trust, fear, and misunderstanding. It is difficult to combat the fear associated with the unknown. Companies and experts in the field must work hard to both gain the trust of the public and communicate what is actually being done with student data. Trust must extend the other way as well. Companies need to trust that by being open about their practices they will not be attacked by concerned external stakeholders. Fear from companies about the reactions of privacy advocates encourages silence on their parts and serves to reduce overall transparency. Both parties must build trust to move towards an open and productive dialogue.

Another recurring theme centers on legislation that has not yet had the desired effect. Privacy advocates view current legislation as too lenient and many companies are able to comply without actually protecting student data. In fact, the legislation may actually harm companies that do the most to protect student privacy. Voluntary pledges offer one solution, though they are not without problems; conflicts of interest often erode even the best self-policing strategies. Many, if not most, companies may support the spirit of such pledges but be unable to sign due to any number various technicalities. Active involvement from all invested parties will be crucial to designing new legislation that will strike a balance between allowing data to be used for the good of education, while protecting the privacy of individual students.

Finally, differing views on the appropriateness of private institutions delivering public goods underscore many of the issues discussed. If commercial vendors are going to be the major providers of educational technologies to school systems there needs to be a shift in how the public perceives these companies. Stifling the success of these companies only serves to keep innovative learning technologies out of the classroom. Still, deference to privacy concerns is an important component of occupying a role in part characterized by public stewardship. Discussions about the ethical limits of financially profiting off of student data need to be addressed directly by corporate, research, and public interests with adequate emphasis on risk and potential system improvements.

Overall, there are variety of issues contributing to concerns over student privacy and how these concerns impact industry applications of educational data mining. These issues are extremely prominent and are not expected to lose momentum soon. The EDM community stands to play an important role in how discussions and legislation around student privacy evolve in the coming years. The landscape of educational data and privacy will continue to shift, and we hope with increased involvement this shift will be positive for researchers and industries interested in using educational data mining to support student learning.

6. REFERENCES

[1] G. Siemens and R. S. J. Baker, "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration," pp. 252–254, 2012.

[2] S. Simon, "Data Mining Your Children," *Politico*, 15-May-2014.

[3] N. Singer, "With Tech Taking Over in Schools, Worries Rise," *The New York Times*, 14-Sep-2014.

[4] S. Trainor, "Student data privacy is cloudy today, clearer tomorrow," *Phi Delta Kappan*, vol. 96, no. 5, pp. 13–18, 2015.

[5] B. Herold, "inBloom to Shut Down Amid Growing Data-Privacy Concerns," *Education Week*, 04-Feb-2014.

[6] "FACT SHEET: Safeguarding American Consumers & Families," *The White House*, 2015. [Online]. Available: http://www.whitehouse.gov/the-press-office/2015/01/12/fact-sheet-safeguarding-american-consumers-families.

[7] J. McQuiggan and A. W. Sapp, *Implement, Improve and Expand Your Statewide Longitudinal Data System: Creating a Culture of Data in Education*. 2014.

[8] Mayer-Schonberger and K. Cukier, *Big Data*. New York, New York: Haughton Mifflin Harcourt Publishing Company, 2013.

[9] U. S. D. of Education, "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief," 2012.

[10] J. Manyika, M. Chui, D. Farrel, S. Van Kuiken, P. Groves, and E. Almasi, "Open data: Unlocking Innovation and Performance with Liquid Information," 2013.

[11] V. Aleven and K. Koedinger, "Limitations of Student Control: Do Students Know When They Need Help?," in *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*, 2000, pp. 292–303.

[12] C. Conati, N. Jaques, and M. Muir, "Understanding attention to adaptive hints in educational games: an eye-tracking study," *Int. J. Artif. Intell. Educ.*, vol. 23, pp. 136–161, 2013.

[13] M. Wen, D. Yang, and C. Rose, "Sentiment Analysis in MOOC Discussion Forums: What does it tell us?," in *Proceedings of the 7th International Conference on Educational Data Mining*, 2014, pp. 257–260.

[14] R. S. J. Baker, A. T. Corbett, and V. Aleven, "More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing," *Knowl. Creat. Diffus. Util.*, pp. 406–415, 2008.

[15] V. Shute, "Stealth Assessment in Computer-Based Games to Support Learning," in *Computer Games and Instruction*, 2011, pp. 503–523.

[16] J. P. Rowe, L. R. Shores, B. W. Mott, and J. C. Lester, "Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments," *Int. J. Artifical Intell. Educ.*, vol. 21, no. 1–2, pp. 115–133, 2011.

[17] J. Sabourin, L. R. Shores, B. W. Mott, and J. C. Lester, "Predicting Student Self-Regulation Strategies in Game-Based Learning Environments," in *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, 2012.

[18] H. Chen, "Identifying Early Indicators for College Readiness," 2007.

[19] L. Pappano, "Using Research to Predict Great Teachers," *Harvard Education Letter*, 2011.

[20] M. Sieber, J. Tolich. "Planning ethically responsible research" Sage Publications, 2012.

[21] S. Carey, "Students, Parents and the School Record Prison A Legal Strategy for Preventing Abuse.pdf," *J. Law Educ.*, vol. 3, p. 365, 1974.

[22] T. L. Elliott, D. Fatemi, and S. Wasan, "Student Privacy Rights — History , Owasso , and FERPA," *J. High. Educ. Theory Pract.*, vol. 14, no. 4, 2014.

[23] J. R. Reidenberg, N. C. Russell, J. Kovnot, T. B. Norton, and R. Cloutier, "Privacy and Cloud Computing in Public Schools," 2013.

[24] R. Silverblatt, "Hiding behind ivory towers: Penalizing schools that improperly invoke student privacy," *Georgetown Law J.*, vol. 101, pp. 493–517, 2013.

[25] B. Smith and J. Mader, "Protecting Students' Privacy - By Law," *Sci. Teach.*, vol. 81, no. December, 2014.

[26] Children's Online Privacy Protection Act of 1998, 5 U.S.C. 6501-6505.

[27] A. Allen, "Minor Distractions: Children, Privacy and E-commerce," Houston Law Review, 2001.

[28] J. Mayfield, "Revised Children's Online Privacy Protection Rule Goes Into Effect Today Federal Trade Commission," *Federal Trade Commission*, 01-Jul-2013.

[29] Data Quality Campaign, "2014 Student Data Privacy Bills," 2014.

[30] E. Brown, "Obama to propose new student privacy legislation," *The Washington Post*, Washington D.C., 19-Jan-2015.

[31] S. Simon, "Barack Obama to seek limits on student data mining," *Politico*, 11-Jan-2015.

[32] H. Tsukayama, "More than 70 companies just signed a pledge to protect student data priacy - with some notable exceptions," *The Washington Post*, 12-Jan-2015.

[33] D. Banisar, "Privacy and data protection around the wolrd," in 21st International Conference on Privacy and Personal Data Protection, 1999.

[34] G. Yee, "Security and Privacy in Distance Education," in Information Security and Ethics: Concepts, Methodologies, Tools, and Applications, 1st ed., H. Namati, Ed. 2007, p. 4110.

[35] D. Boyd, "Response to COPPA Rule Review, 16 CFR part 312, Project No. P-104503," Washington D.C., 2011.

[36] N. S. B. Association, "Data in the Cloud: A Legal and Policy Guide for School Boards on Student Data Privacy in the Cloud Computing Era," Alexandria, VA, 2014.

[37] C. S. Media, "Student Privacy Survey," 2014.

[38] S. Fox, "In the Matter of COPPA Rule Review, 16 CFR Part 312, Project No. P-104503," Washington D.C., 2011.

[39] J. Palfrey, D. Boyd, and U. Gasser, "How the COPPA, as Implemented, Is Misinterpreted by the Public: A Research Perspective," 2010.

[40] Pew Research Center, "What Internet Users Know about Technology and the Web," 2014.

[41] C. DeLorme, "Response to COPPA Rule: Comments to be placed on the public record," Washington D.C., 2012.

[42] M. Madden, S. Cortesi, U. Gasser, A. Lenhart, and M. Duggan, "Parents, Teens, and Online Privacy," 2012.

[43] B. L. Wilcox, D. Kunkel, J. Cantor, P. Dowrick, S. Linn, and E. Palmer, "Report of the APA Task Force on Advertising and Children," 2004.

[44] K. Vanlehn, "The Behavior of Tutoring Systems," *Int. J. Artif. Intell. Educ.*, vol. 16, no. 3, pp. 227–265, 2006.